Analysis of a family 6 Carbohydrate-Binding Module and three family 69 Hyaluronidases

Nicola Louise Smith

A thesis submitted in partial fulfilment
of the requirements of the University of
Northumbria for the degree of Doctor of Philosophy

This research program was carried
out in collaboration with York Structural
Biology Laboratory at the University of York
and Newcastle University.

May 2004

Table of Contents

Table of Figures

Table of tables

Abstract

To investigate the interactions between a family 6 carbohydrate-binding module (CBM) from *Clostridium thermocellum* Xyn11A (*Ct*CBM6) and its target ligands and to identify the location of the ligand binding site(s) through a mutagenesis strategy, the protein was expressed in *Escherichia coli* and purified to homogeneity. *Ct*CBM6 was shown previously to interact with xylan (Fernandes *et al.*, 1999) and, informed by the crystal structure, it was found that *Ct*CBM6 was unusual, as it contained two potential ligand-binding clefts (Clefts A and B).

Qualitative ligand specificity studies through affinity gel electrophoresis (AGE) demonstrated that *Ct*CBM6 bound preferentially to xylans, interacts weakly with β-glucan and some soluble substituted forms of cellulose. Quantitative analysis of ligand binding by isothermal titration calorimetry showed that *Ct*CBM6 bound xylooligosaccharides from xylobiose to xylohexaose, with affinity increasing with chain length. The affinity of *Ct*CBM6 for soluble xylan of varying degrees of substitution was judged to be similar.

NMR spectroscopy (Dr M. Czjzek at CNRS, Marseille) indicated that xylohexaose interacts with the two solvent exposed aromatic amino acids (Tyr-34 and Trp-92) and a polar amino acid (namely Asn-120) in cleft A of *Ct*CBM6. Site-directed mutagenesis revealed that hydrophobic stacking interactions and hydrogen bonds potentiate the binding of *Ct*CBM6 to xylan. Surface aromatic residues Tyr-34 and Trp-92 of *Ct*CBM6 are pivotal in the interaction between this module and its ligand, as substitution of these amino acids with alanine and methionine resulted in an 8-fold and 50-fold respective decrease in affinity of *Ct*CBM6 for oat spelt xylan, as judged by quantitative AGE. Hydrogen-bonding interactions also made pivotal contributions to the overall binding in *Ct*CBM6. Asn-120 was critical to ligand binding, as the mutant N120A showed ~130-fold loss of binding affinity. This suggests that this residue directly participates in ligand binding via hydrogen bonds. Collectively, mutagenesis and NMR studies showed that cleft A can accommodate xylooligosaccharides and xylan, while cleft B was unable to interact with target ligands.

Three hyaluronidases (HylP1, HylP2 and HylP3) of glycoside hydrolase family 69 (GH69) were cloned from the genome sequenced organism *Streptococcus pyogenes* SF370, expressed in *E. coli* and purified to homogeneity. Characterisation of the N-terminally tagged HylP1 (38.4 kDa), HylP2 (42.0 kDa) and HylP3 (41.8 kDa) revealed activity against sodium hyaluronate with a K_M for HylP1, HylP2 and HylP3 of 0.90, 2.07 and 4.35 ml mg^{-1}, and a k_{cat} of 1390.90, 742.01 and 1253.04 s^{-1}, respectively. HylP1, HylP2 and HylP3 displayed an optimum pH of 6.5, 6.0 and 5.5, respectively, and an optimum activity at 37 °C. Moreover, PAGE analysis showed each enzyme was endo-cleaving.

All three enzymes have been crystallised and sufficient quality diffraction data obtained for HylP1 and HylP3 (data collection and processing was performed by Dr Edward Taylor). The 3D structure of HylP1 has been solved at a resolution of 1.8 Å and is composed of three monomeric strands that are intertwined to form a trimer.

Abbreviations

A	Adenosine (DNA) / Alanine (protein)
Å	Angstrom
A_x	Absorbance at x nm
AGE	Affinity gel electrophoresis
Amp^r	Ampicillin resistence
APS	Ammonium persulphate
ATP	Adenosine triphosphate
bp	Base pair(s)
BPB	Bromophenol blue
BSA	Bovine serum albumin
C	Cytosine
°C	Degree Celsius
CBD	Cellulose-binding domain
CBM	Carbohydrate-binding module
CBM6	Carbohydrate-binding module family 6
CE	Carbohydrate esterase
CFE	Cell free extract
CMC	Carboxymethylcellulose
Cm^r	Chloramphenicol resistence
CS	Chondroitin sulphate
CSS	Clear strategy screen
C-terminal	Carboxy terminal
*Ct*CBM6	*Clostridium thermocellum* CBM6
*Ct*Xyn11A	*Clostridium thermocellum* xylanase 11A
3D	Three-dimensional
Da	Dalton
DMSO	Dimethyl sulphoxide
DNA	Deoxyribonucleic acid
dsDNA	Double stranded DNA
DNSA	Dinitrosalicyclic acid
dNTP	Deoxynucleotriphosphate
DS	Dermatan sulphate
DTT	Dithiothreitol
ECM	Extracellular matrix
EDTA	Ethylene diamine tetraacetic acid, disodium salt
EtBr	Ethidium bromide
F	Phenylalanine
g	Gram(s)
G	Guanidine
ΔG^o	Change in free energy
G_5	Cellopentaose
G_6	Cellohexaose
GAG	Glycosaminoglycan
GAS	Group A streptococci
GluA	Glucuronic acid
GH	Glycoside hydrolase
GH69	Glycoside hydrolase family 69
GT	Glycosyltransferase

h	Hour(s)
ΔH^{o}	Change in enthalpy
HA	Hyaluronan/Hyaluronic acid
HEC	Hydroxyethyl cellulose
HEPES	N-[2-Hydroxyethyl]piperazine-N'-[2-ethanesulphonic acid]
HEWL	Hen egg white lysozyme
His_6 tag	Hexahistidine tag
HPAEC	High performance anion exchange chromatography
HS	Heparan sulphate
HSQC	Heteronuclear single quantum coherence
i.d.	Internal diameter
IMAC	Immobilised metal affinity chromatography
IPTG	Isopropyl-β-D-Thiogalactopyranoside
ITC	Isothermal titration calorimetry
K	Kelvin
K_a	Association constant
Kan^r	Kanamycin resistance
kb	Kilobase pair(s)
k_{cat}	Turnover number
k_{cat}/ K_M	Catalytic efficiency
K_d	Dissociation constant
kDa	Kilodalton
K_M	Michaelis-Menten constant
KS	Keratan sulphate
L	Litre(s)
LB	Luria-Bertani medium
m	Metre(s)
M	Molar / Methionine
mA	Milliamps
MAD	Multiple anomalous difference
MES	Morpholinoethanesulfonic acid
min	Minute(s)
mm	Millimetre(s)
MME	Monomethylether
MOPS	4-(N-morpholino)propanesulphonic acid
MPD	2-methyl-2,4-pentanediol
M_r	Relative molecular mass
MW	Molecular weight
N	Asparagine
NAG	N-acetylglucosamine
NMR	Nuclear magnetic resonance
nt	nucleotide(s)
N-terminal	Amino terminal
OD_x	Optical density at x nm
ORF	Open reading frame
p	Plasmid
p	Plasmid DNA
PAGE	Polyacrylamide gel electrophoresis
PCR	Polymerase chain reaction
PCW	Plant cell wall

PEG	Polyethylene glycol
PL	Polysaccharide lyase
PPF	Periplasmic fraction
Psi	Pounds per square inch
R	Arginine
rpm	Revolutions per minute
RT	Room temperature
s	Second(s)
ΔS°	Change in entropy
SAD	Single anomalous difference
SDM	Site-directed mutagenesis
SDS	Sodium dodecylsulphate
Se-Met	Seleno-methionine
SLO	Streptolysin O
SLS	Streptolysin S
sp.	Species
Spe	Streptococcal pyrogenic toxin
subsp.	Subspecies
ssDNA	Single stranded DNA
T	Temperature
TAE	Tris-Acetate-EDTA
TBE	Tris-Borate-EDTA
TEMED	N,N,N',N'-tetramethylethylene diamine
Tet^r	Tetracycline resistance
T7 *lac*	*lac* operator just downstream of T7 promotor
T_m	Melting temperature
Tris	tris(hydroxymethyl)aminomethane
$T\Delta S^{\circ}$	Change in entropy
TSST	Toxic shock syndrome toxin
UDP	Uridine diphosphate
UV	Ultraviolet
V	Volt(s)
V_{max}	Maximum velocity
v/v	Volume per volume
W	Trytophan
w/v	Weight per volume
X_1	Xylose
X_2	Xylobiose
X_3	Xylotriose
X_4	Xylotetraose
X_5	Xylopentaose
X_6	Xylohexaose
XBD	Xylan-binding domain
x g	Times gravity
X-gal	5-bromo-4-chloro-3-indoyl-β-D-galactoside
Y	Tyrosine
ε	Molar absorptivity
α	Alpha
β	Beta
Δ	Delta

k	Kilo
λ	Lambda
m	milli
μ	Micro
n	Nano
θ	Oscillation angle
p	Pico
Ψ	Oscillation range
1°	Primary
2°	Secondary
3'	Three prime
5'	Five prime
18.2 MΩ H_20	18.2 mega ohm water

1 General introduction

In the simplest sense, cell membranes are a carbohydrate and lipoprotein structure that surround and separate cells (Kohorn, 2000). Every living cell contains a membrane, often referred to as the plasma membrane. In addition, many plant and bacterial cells contain a cell wall that surrounds the plasma membrane. The plant cell wall (PCW), which is an integral part of the cell (Wyatt & Carpita, 1993), can be compared to the animal extracellular matrix (ECM) because it shows functional similarities (Kohorn 2000; Wojtaszek, 2000). Polysaccharides are integral components of living organisms. Plant cells contain cellulose, xylan, xyloglucan, mannan, glucomannan, arabinan, homogalacturonan, rhamnogalacturonan, galactan and arabinogalactan and animal cells contain chondroitin sulphate, dermatan sulphate, heparan sulphate, heparin, keratan sulphate and hyaluronan, which are also found in microbes, such as *Streptococcus pyogenes*. This thesis analyses, at the molecular level, proteins that interact with polysaccharides, specifically a carbohydrate-binding module (CBM) that interacts with the plant polysaccharide xylan, and three enzymes that hydrolyse the animal and microbial polysaccharide hyaluronan.

This introduction will firstly describe the environment in which these polysaccharides are found, i.e. the PCW, in the case of xylan, and the animal connective tissue and microbial capsule, in the case of hyaluronan; the structures of the polysaccharide and then the proteins that interact with these polysaccharides.

1.1 Environments in which polysaccharides are found

1.1.1 The plant cell wall

Approximately 4 x 10^{10} tonnes of cellulose alone are synthesised annually as a result of photosynthesis (Coughlan, 1985), although not all this material accumulates because bacteria and fungi enzymically degrade PCWs to provide a source of carbon and energy, thus playing a major role in recycling organic carbon in the biosphere. A very important group of carbohydrates is the structural polysaccharides located in PCWs, as these polymers are the most abundant

organic molecules in the biosphere (Warren, 1996; Simpson *et al.*, 2000). A lack of basic knowledge of the process of PCW hydrolysis has driven much of the initial work into PCW hydrolases, and the role CBMs play within these enzymes is of great interest.

Plant cells, unlike animal cells, possess a highly structured extracytoplasmic compartment known as the cell wall. The PCW has a number of important functions in the plant. This semi-permeable wall is the first barrier encountered by pathogens and other organisms; it provides mechanical support to the cell as well as giving plant cells the capacity to withstand osmotic pressure, it physically controls the rate of cell expansion, provides a store of food reserves and controls intercellular transport (Brett & Waldron, 1996; Carpita, 1997). This plethora of functions is a possible reason for the complex structure of the PCW.

The plant cell deposits its wall as a series of layers. There are two basic types of PCW, the thin primary wall that is laid down by young and undifferentiated cells, and the secondary walls that are thicker and are synthesised after the cell has stopped growing (Albersheim *et al.*, 1994; Cosgrove, 1997). The composition of the primary cell wall comprises cellulose, hemicellulose, pectin and glycoproteins. The secondary cell wall, which arises from the increase in cell wall thickness in mature plants, is composed mainly of lignin. In addition, the cementing middle lamella, which is the connective tissue between PCWs, and is sandwiched between the primary wall of each adjoining cell, contains mainly pectic polysaccharides and calcium salts (Brett and Waldren, 1996). The structurally important components in the PCW are polysaccharides, which are typically cellulose, hemicelluloses and pectins, although lignin, present in the mature plant, can also play a key role in the structural integrity of PCWs (Carpita, 1997).

1.1.1.1 The structure of the plant cell wall

The general structure of both the primary and secondary wall layers is similar and consists of highly crystalline cellulose microfibrils embedded in an amorphous PCW matrix composed mainly of a variety of different polymer networks of hemicelluloses, pectic polysaccharides, with some glycoproteins and lignins (usually only in secondary walls) also present, which interact together giving rise

to the structure as a whole (Brett and Waldren, 1996; Carpita, 1997). A model of the PCW is shown in Figure 1.1.

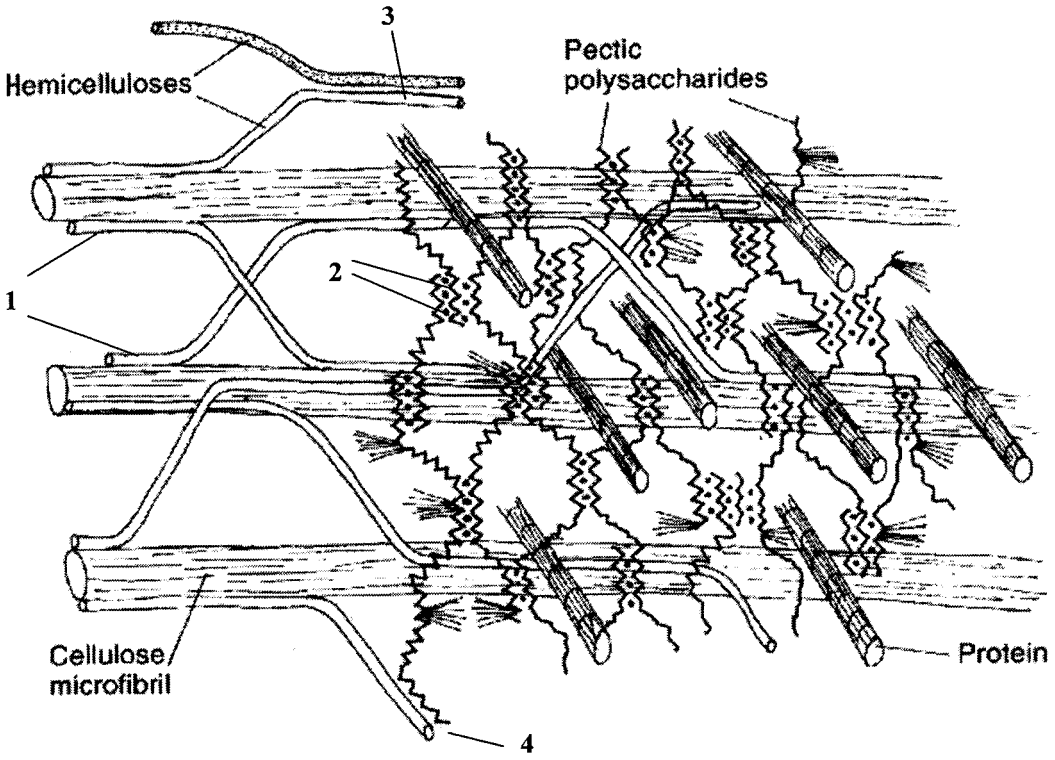


Figure 1.1: A generalised model of the PCW.
Possible interactions between several classes of cell wall polymer and the likely orientation: (1) hydrogen bonding between cellulose and hemicellulose; (2) ionic bonding between galacturonic acid residues of pectin (calcium ions act as ionic bridges); (3) hydrogen bonding between hemicelluloses; (4) hydrogen bonding between hemicellulose and pectin. Modified from Brett and Waldren (1996).

1.1.2 Animal connective tissue

Connective tissue is the most abundant and widely distributed tissue in the body (Thibodeau and Patton, 1992). It is found in skin, membranes, muscles, bones, nerves and all internal organs. The functions of connective tissues are as varied as its structure and appearance. Connective tissues are responsible for the form and shape of the animal body and, in addition, provide protection for vital organs and facilitate locomotion. Connective tissue exists as delicate webs that hold internal organs together and give them shape. It also exists as strong and tough cords, rigid bones and even in the form of a fluid (i.e. blood; Thibodeau and Patton, 1992).

In vertebrates, the most common type of connective tissue is loose connective tissue. It holds organs in place and attaches epithelial tissue to other underlying tissues. Loose connective tissue is named based on the 'weave' and type of its constituent fibres. There are three main types, namely: collagenous fibres, which are made of collagen; elastic fibres, which are composed of elastin; and reticular fibres, which join connective tissue to other tissues. Another type of connective tissue is fibrous connective tissue that is found in tendons and ligaments (Kuettner, 1992). It is composed of large amounts of closely packed collagenous fibres. Specialised connective tissues include adipose tissue, which is a form of loose connective tissue that stores fat. Areolar connective tissue is the 'glue' that gives form to the internal organs. Bone is a type of mineralised connective tissue that contains collagen and calcium phosphate giving bone its firmness. Blood is also considered to be a type of connective tissue. Even though it has a different function in comparison to other connective tissues, it does have an ECM. The matrix is the plasma and erythrocytes, leukocytes and platelets are suspended in the plasma. Finally, cartilage is a form of fibrous connective tissue that is composed of closely packed collagenous fibres.

Articular cartilage is the connective tissue that covers ends of long bones in synovial joints providing smooth articulation during joint movement (Knudson and Knudson, 2001). Normal articular cartilage is unique within the class of connective tissue in that it lacks blood vessels, lymphatic vessels, nerve fibres and has no separating basement membrane on either side of the tissue (Kuettner,

1992). Cartilage consists of two principle components: the tissue fluid and the framework of structural macromolecules that give the tissue its form and stability. The structural macromolecules of the cartilage: collagen, proteoglycans and non-collagenous proteins contribute 20-40 % of the wet weight of the tissue (Buckwalter and Mankin, 1998).

The specific structural organisation of articular cartilage matrix gives it biomechanical properties such as tensile strength and elasticity, which enable it to absorb and distribute loads (Huber *et al.*, 2000). The organisation of collagen fibres into an ordered three-dimensional (3D) network provides tensile strength and is essential for maintaining the tissues volume and shape. The highly hydrated negatively charged proteoglycans give the cartilage its ability to undergo reversible deformation and resist compressive forces. Proteoglycans are crucial in distributing load in weight-bearing joints (Hardingham and Fosang, 1992). The proteoglycans attract water and provide a swelling pressure. Under normal conditions, the stiff collagen network limits the swelling of the proteoglycans (Kuettner and Thonar, 1998). Upon loading, cartilage matrix is deformed by the expulsion of fluid from the loaded region, with movement of water towards the synovial cavity and away from the proteoglycan rich region. The water carries away cellular waste products. Proteoglycans are forced closer together, increasing the negative charge density and the intermolecular charge repulsive forces, which in turn increase the resistance of the tissue to further deformation. Finally, the deformation reaches an equilibrium in which the external loading force is balanced by the internal forces based upon the swelling pressure. Once the load is removed, water returns from the synovial cavity, carrying with it nutrients for the cells.

The pericellular environment is characterised by a high content of large proteoglycan aggregates. Chondrocytes, the only cell type found in articular cartilage, have CD44-like surface receptors, which interact with hyaluronic acid (Knudson *et al.*, 1993), thus binding proteoglycans to the cell. The pericellular environment contains very little organised fibrillar collagens, but a high concentration of filamentous type VI collagen that also binds to hyaluronic acid and interacts with the cell surface (Knudson and Knudson, 2001).

1.1.2.1 The structure of connective tissue

Connective tissue differs from epithelial tissue in the arrangement and variety of its cells and in the amount and kinds of intercellular material, called matrix, found between its cells (Thibodeau and Patton, 1992). Besides the relatively few cells embedded in the matrix of most types of connective tissue, varying numbers and kinds of fibres are also present (Figure 1.2). The structural quality and appearance of the matrix and fibres determine the qualities of each type of connective tissue (Knudson and Knudson, 2001). All cells of the body, however, make contacts with surrounding structures that involve connective tissues as components of the ECM. The matrix possesses chemical, physical, and mechanical properties uniquely suited to the function of tissues and organs of which the cells are a part. The major components of the ECM are fibrillar proteins (collagens and elastin), globular proteins, and, in the case of bone, the inorganic mineral phase. Other major components of connective tissues include the high molecular weight carbohydrates that are composed of a glycosaminoglycan (GAG) portion (the complex carbohydrate itself) linked to a core protein. The core protein with the glycosaminoglycans attached is termed the proteoglycan subunit. Another abundant complex carbohydrate present in many tissues and the major polysaccharide of synovial fluid is hyaluronic acid (HA). The complex carbohydrates of cartilage are composed of high molecular weight polymers of the proteoglycan subunits, with the polysaccharide side chains of chondroitin sulfate and keratan sulfate linked to serine residues of the core protein. The polymeric components consist of these proteoglycan subunits bound to high molecular weight HA chains through interactions with another glycoprotein, called link protein. These proteoglycan aggregates are envisioned as occupying the spaces in cartilage surrounded by the collagen fibers and other components of the matrix. Both the proteoglycans and the collagens are synthesized by the articular chondrocytes (Muir, 1995; Buckwalter and Mankin, 1997).

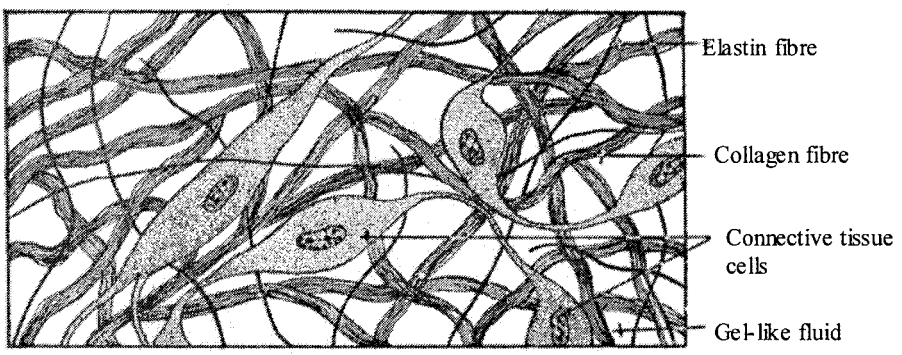


Figure 1.2: Normal connective tissue.
Connective tissue forms the glue between cells in the body's tissues, holding the cells together and giving those tissues structure and shape. It is made of a number of different molecules, including a range of fibres that give strength and elasticity to many tissues. Taken from http://www.marfan.net.au/.

1.1.3 The microbial capsule

Bacteria have a variety of structures outside the cell wall that function in protection, attachment or cell movement. For example, some bacteria possess a well organised layer of material lying outside the cell wall known as a capsule. The major human host defence against invasive group A streptococcal infection is that of phagocytosis and killing by polymorphonuclear leucocytes (Bisno *et al.*, 2003). Thus, a critical group A streptococcal virulence factor is an antiphagocytic capsule composed of hyaluronic acid (Wessels and Bronze, 1994). Streptococcal strains vary greatly in their degree of encapsulation and those with the most exuberant capsule production have a mucoid appearance (Bisno *et al.*, 2003). The association of encapsulation with resistance to phagocytosis has been recognised for many years (Wessels *et al.*, 1991; Wessels and Bronze, 1994; Dale *et al.*, 1996; Cunningham, 2000; Bisno *et al.*, 2003). Wessels and Bronze (1994) showed that acapsular mutant strains were altered in their ability to resist phagocytic killing and thus its mouse virulence decresed significantly. In addition, the capsule may be an important adherence factor in the pharynx, since it binds CD44 on epithelial cells (Schrager *et al.*, 1998).

1.1.3.1 The structure of the microbial capsule

Group A streptococci are covered with an outer hyaluronic acid capsule (Stoolmiller and Dorfman, 1969; Cunningham, 2000), while the group A carbohydrate antigen and the type-specific M protein are attached to the bacterial cell wall and membrance, as shown in Figure 1.3. Both the M protein and the capsule are considered virulence factors conferring antiphagocytic properties upon the streptococcal cell (Foley and Wood, 1959; Moses *et al.*, 1997; Cunningham, 2000). Synthesis of the polymer of hyaluronic acid for capsule expression in group A streptococci is controlled by an operon composed of three different genes: *hasA*, *hasB* and *hasC* (Crater and van de Rijn, 1995). *HasA* encodes hyaluronate synthase, *hasB* codes for uridine diphosphate (UDP)-glucose dehydrogenase and *hasC* encodes UDP-glucose pyrophosphorylase (Alberti *et al.*, 1998). However, inactivation of the *hasC* gene does not affect capsule production, suggesting that an alternative source of UDP-glucose is available to the bacteria for capsule production (Asbaugh *et al.*, 1998).

Group A streptococcal strains rich in M protein and capsule are extremely virulent for humans, and mucoid strains have long since been see to cause deeply invasive infections (Moses *et al.*, 1997). Group A streptococcal capsular hyaluronate is chemically quite similar to that found in human connective tissue (Bisno *et al.*, 2003). For this reason it is a poor immunogen and antibodies to group A streptococcal hyaluronic acid have been quite difficult to demonstrate in humans.

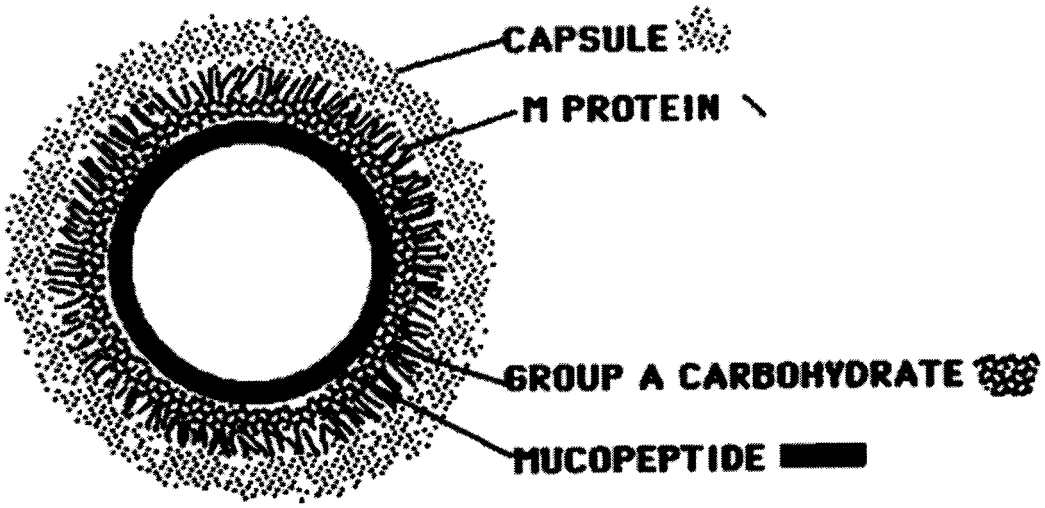


Figure 1.3: Diagram of the group A streptococcal cell.
Group A streptococcal cell is covered with an outer hyaluronic acid capsule and the group A carbohydrate, consisting of a polymer of rhamnose with *N*-acetylglucosamine side chains. Streptococcal M protein extends from the cell wall and is anchored in the membrane. Taken from Cunningham, 2000.

1.2 Biochemistry and structure of specific polysaccharides

1.2.1 Cellulose

Cellulose is the most abundant polysaccharide on earth and this linear homopolymer is a β-1,4-linked polymer of D-glucose residues. The repeating unit is cellobiose and each glucose residue is rotated 180 ° about the main chain axis with respect to its neighbouring residues, so all the sugar rings lie in the same plane (Béguin and Aubert, 1994). The average chain length within the cell wall has been estimated at 100 – 1400 saccharide residues (Béguin and Aubert, 1994). As the polymer is unsubstituted and planar, adjacent strands align parallel to the long axis to form cellulose microfibrils. The chains are held together in layered sheets by an extensive network of inter- and intramolecular hydrogen bonds and van der Waals forces. These microfibrils are disorganised on the surface, but their regular interior is highly crystalline in form, resulting in high tensile strength (Coughlan, 1985). Although cellulose is highly crystalline, there is evidence to show that some degree of structural variation exists within the microfibril, consisting of amorphous or paracrystalline regions (Teeri, 1997). Disorganised bonding interactions between adjacent strands results in the generation of amorphous regions, which mainly occur near the surfaces but can also span the entire width of the crystalline cellulose, and may be caused by twisting and straining of the microfibril (Teeri, 1997). These amorphous regions permit hemicellulosic polymers to become intimately associated with the cellulose.

1.2.2 Hemicellulose

Hemicelluloses are the major component of the PCW matrix. Hemicellulose is the major non-cellulosic components in most plant cells and contain polymers rich in glucose, xylose, or arabinose that, unlike cellulose, have extensive side chains often including xylose, galactose, and fucose (Kohorn, 2000). They are often found in tight association with the cellulose microfibrils, due to the formation of hydrogen bonds between the two polysaccharides. Hemicelluloses are also found spanning the gaps between microfibrils, forming a network that holds the microfibrils in place, which provides structural strength to the wall (Brett and

Waldron, 1996). Hemicelluloses are chemically more complex than cellulose and often contain various different side chains. The presence of the side chains and/or the nature of the main chain in most hemicelluloses prevent the tight packing of the molecules into the highly ordered crystalline lattices like cellulose. Instead, they tend to be less ordered open structures. The four predominant forms of hemicellulose are classified according to the composition of their sugar backbone and include xylans, xyloglucans, mannans, and glucomannans. In most cell types, xylan or xyloglucan predominates, with others present in smaller amounts (Brett and Waldren, 1996).

1.2.2.1 Xylan

Xylans are the major hemicellulosic polysaccharides present in the cell walls of many plants. Xylan has a main chain composed of β-1,4-linked xylose residues and are usually found substituted with either acetyl, arabinosyl and glucuronosyl residues (Gilbert and Hazelwood, 1993). As xylose differs from glucose by the absence of a C-5 hydroxymethyl group, main chain xylan has a less extensive intramolecular hydrogen bonding network compared to cellulose. Also, adjacent xylan backbones are prevented from forming significant intermolecular bonding interactions because of their generally highly substituted nature, as illustrated in Figure 1.8. The chain length and degree and type of substitution are dependent on the origin of the xylan. Birchwood xylan possesses very little substitution except for acetyl groups, and xylans from oat spelt contain about 20 % arabinose units and some glucose and galactose molecules (Li *et al.*, 2000).

Xylan plays a crucial role in cross-linking the various PCW polysaccharides. This process is mediated by covalent bonds between the arabinosyl side chains of arabinoxylans and the phenolic components of lignin (Sunna and Antranikian, 1997).

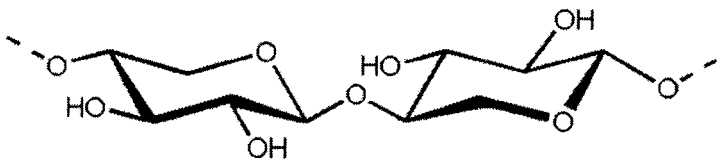


Figure 1.4: Structure of Xylobiose.
Xylan is composed of β-1,4-linked xylose residues. Taken from Pires *et al.*, 2004.

1.2.2.2 Xyloglucan

Xyloglucans are composed of a backbone of β-1,4-linked glucose residues, to the majority of which (~ 75 %) xylose residues are attached via α-1,6-bonds (Carpita, 1997). Some xylose side chains are further substituted at the C-2 position with α-linked arabinose, or the α-linked disaccharide fucopyranosyl-β-1,2-linked to galactose. Acetyl groups may also be present on a high proportion of galactose residues. Xyloglucan in some tissues is known to form a fairly regular structure of nonasaccharide and heptasaccharide units (Brett and Waldren, 1996). Xyloglucans can bind tightly to the surface of cellulose microfibrils via multiple hydrogen bonds (Gibeaut and Carpita, 1994). This cross-linking of cellulose microfibrils is thought to give strength to the wall as well as having an important regulatory role in limiting the rate of elongation growth (Albersheim *et al.*, 1994).

1.2.2.3 Mannans

Mannans are a group of hemicelluloses that include a linear homopolysaccharide of β-1,4-linked mannose residues and galactomannan, which consists exclusively of a β-1,4-linked mannan backbone with side chains of α-1,6-linked galactose residues added at the O-6 position (Carpita, 1997).

1.2.2.4 Glucomannans

Glucomannans consist of a backbone of β-1,4-linked glucose and mannose residues without any regularity in sequence. The mannosyl units of the backbone are often substituted with galactose at C-6 and O-acetyl groups at C-2 or C-3 positions (Brett and Waldren, 1996). These α-1,6,-linked galactose substitutions result in the hemicellulose galactoglucomannans.

1.2.3 Pectin polysaccharides (Pectins)

These are a group of complex polysaccharides rich in galacturonic acid, rhamnose, arabinose and galactose and include rhamnogalacturonan, arabinan, galactan and arabinogalactan. The pectic polymers form a network that is largely structurally independent of the cellulose/hemicellulose network and may form a jelly-like matrix (Carpita and Gibeaut, 1993; Kohorn, 2000). They are thought to perform many functions such as providing physical strength at cell junctions,

determining wall porosity and providing charged surfaces that modulate cell wall pH and ion balance (Brett and Waldren, 1996).

1.2.3.1 Rhamnogalacturonan/homogalacturonan composite polysaccharide

This polysaccharide contains a backbone of α-1,4-linked galacturonic acid and α-1,2-linked rhamnose, to which the side chains of arabinose and galactose attach to the C4 position of rhamnose. This composite polymer consists of two discrete domains, comprising regions of polygalacturonic acid (homogalacturonan) of varying methyl esterification (smooth regions), and rhamnogalacturonan, which has arabinan and galactan rich side chains attached to about half of the rhamnose units (hairy regions) (Varner and Lin, 1989; Brett and Waldren, 1996).

1.2.3.2 Arabinan

Arabinan is a highly branched molecule consisting of a backbone of α-1,5-linked arabinose residues that are frequently substituted with either monomeric or oligomeric α-1,2- and/or α-1,3-linked arabinose side chains to form the branch points. In the PCW, arabinan is usually associated with pectin, via an α-1,4-linkage to the rhamnose residues of rhamnogalacturonan (Lerouge *et al.*, 1993).

1.2.3.3 Galactan and arabinogalactan

Galactan consists of primarily a β-1,4-linked galactose backbone, to which short α-1,5-linked arabinose side chains are attached at the C3 position of the main chain to form arabinogalactan (Brett and Waldren, 1996). They are found as either independent molecules or as side chains attached to rhamnogalacturonan.

1.2.4 Glycoproteins

As well as polysaccharides, cell walls also contain a variety of different proteins, most of which are glycosylated. One of the most abundant glycoproteins is extensin. Extensin is a hydroxyproline rich protein that forms an extensive insoluble network that is laid down perpendicular to the cellulose/hemicellulose network. It is thought that extensins hold this network together by the formation of covalent and non-covalent bonds (Carpita and Gibeaut, 1993).

1.2.5 Glycosaminoglycans and proteoglycans

Proteoglycans consist of one or more glycosaminoglycan (GAG) chains, which are linked covalently to a protein core. The protein core of different proteoglycans can take many forms. The GAG component includes one or more of chondroitin sulphate (CS), dermatan sulphate (DS), heparan sulphate (HS), heparin and keratan sulphate (KS). All tissues contain proteoglycans, either at the cell surface of cells or within the extracellular space. Proteoglycans are most abundant in the connective tissues, especially in articular cartilage. The largest in size and most abundant proteoglycan in articular cartilage is aggrecan, representing about 90 % of the total cartilage proteoglycan mass. The central component of the cartilage proteoglycan aggregate is a long molecule of hyaluronan (Figure 1.5). Aggrecan is characterised by its ability to interact with hyaluronic acid to form large hydrated proteoglycan aggregates, which are decorated with GAGs (Roughley and Lee, 1994).

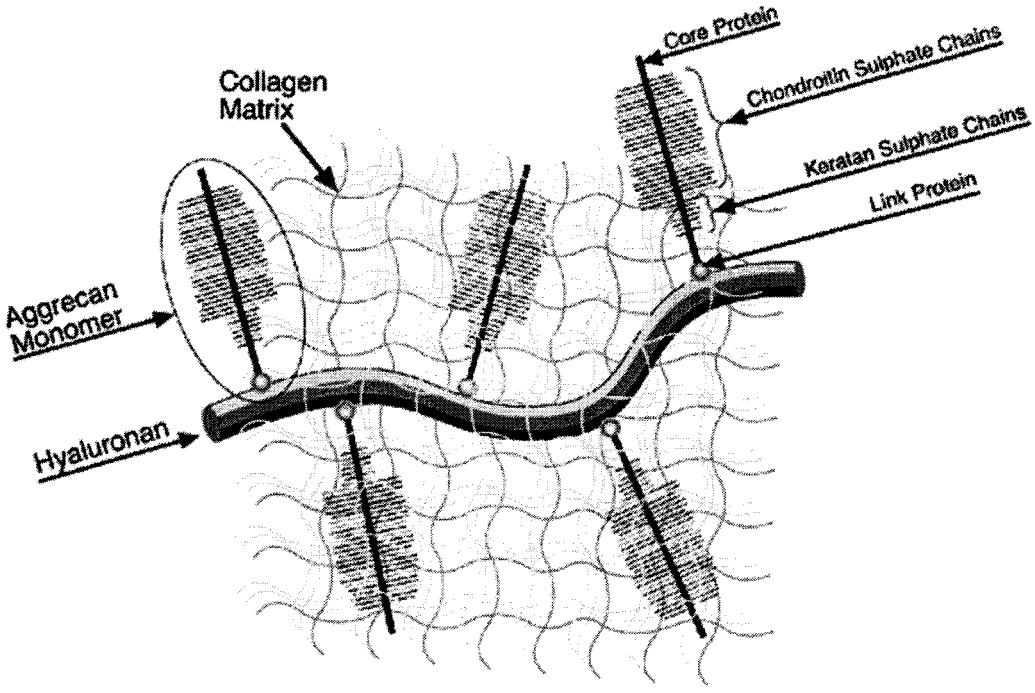


Figure 1.5: Schematic representation of the structure of cartilage proteoglycan aggregate.
The N-terminal domain of the core protein binds to a hyaluronan molecule. Binding is facilitated by a link protein, which binds to both hyaluronan and the aggrecan core protein. Each aggrecan core protein has multiple chains of chondroitin sulphate and keratan sulphate covalently attached. Taken from http://www.peprotech.com/content/focusarticles.htm?id=72.

Aggrecan is a large molecule with a relative molecular mass (M_r) of 2000-3000 kDa. Aggrecan consists of a protein core (230 kDa) with three globular domains. One of the globular domains of the aggrecan core protein has an N-terminal that binds with high affinity to a decasaccharide sequence in hyaluronan; this binding is facilitated by a link protein that binds to the aggrecan core protein and hyaluronan (Figure 1.5) (Zubay, 1998). During post-translational processing a large number of covalently attached GAG side chains and *N*-linked and *O*-linked oligosaccharides are added (Knudson *et al.*, 1993). These GAG chains are CS and KS, long unbranched polysaccharide chains with highly charged sulphate and carboxylate groups. GAGs form long strings of negative charges that attract counter ions and the osmotic balance caused by a local high concentration of ions draws water from the surrounding areas. Consequently, proteoglycans keep the matrix hydrated.

The GAG components of proteoglycans are composed of alternating sequences of hexosamine and either D-glucuronic acid or L-iduronic acid. GAGs can be classified by their hexosamine components; D-glucosamine-containing GAGs are known as glucosaminoglycans and include HS. CS and DS are known as galactosaminoglycans because D-galactosamine is present (Lindahl *et al.*, 1994). Hyaluronan (or hyaluronic acid; HA), on the other hand, is unique amongst GAGs as it is not covalently bonded to proteoglycan, and it is the only naturally occurring GAG that is not sulphated (Frost *et al.*, 1996; Csóka *et al.*, 1997). A summary of the composition and relative sizes of the different GAGs is shown in Table 1.1. All the members of this family are sulphated and covalently linked to proteins except for HA. There is only one primary conformation of HA unlike other GAGs, which can have almost infinitely variable primary conformations, through the post-synthetic addition of sulphate groups (Laurent *et al.*, 1992).

GAG	Monosaccharides		Sulphate per disaccharide	Molecular weight (kDa)
Chondroitin sulphate	Glucuronic acid	*N*-acetyl galactosamine	0.2 – 2.3	5 – 50
Dermatan sulphate	Glucuronic acid	*N*-acetyl galactosamine	1.0 – 2.0	14 – 40
Heparan sulphate	Glucuronic acid	*N*-acetyl glucosamine	0.2 – 2.0	5 – 12
Heparin	Glucuronic acid	*N*-acetyl glucosamine	2.0 – 3.0	6 – 25
Hyaluronic acid	Glucuronic acid	*N*-acetyl glucosamine	0	4 – 800
Keratan sulphate	Galactose	*N*-acetyl glucosamine	0.9 – 1.8	4 - 19

Table 1.1: Composition of glycosaminoglycans found within the extracellular matrix.
Based upon data from Alberts *et al.* (1989), Laurent *et al.* (1992) and Knudson *et al.* (1993).

1.2.5.1 Hyaluronan

HA was first isolated by Karl Meyer in 1934 from the vitreous humour of the eye. For many years it was considered just a space filling molecule in the ECM. This idea has changed with the discovery of cell surface binding receptors, suggesting a more complex biological role (Laurent *et al.*, 1992). HA has now been implicated in many different cellular processes including cell adhesion, aggregation, proliferation, inflammation and locomotion. Due to its involvement in so many diverse processes, HA can play a major role in both normal and abnormal functions of cells (Knudson *et al.*, 1993).

HA is a high molecular weight polysaccharide (4-800 kDa), a repeating disaccharide straight chain polymer with the structure (*N*-acetyl-D-glucosamine-D-glucuronic acid)$_n$ (Scott, 1995; Frost *et al.*, 1996; Csóka *et al.*, 1997; Figure 1.6). HA carries a high net negative charge from the carboxyl groups of the glucuronic acid residues, which in biological systems are associated with many mobile cations, such as Na^+, K^+, Ca^{2+} and Mg^{2+}, forming a neutral salt (Laurent *et al.*, 1992).

Studies of HA structure have revealed that it exists as a single, long, unbranched and unsulphated linear polysaccharide chain. It has been demonstrated that each disaccharide unit is rotated through 180° with respect to the previous unit. This rotation means that every two disaccharide units, the HA chain rotates through 360°, to form a twisted structure known as a two-fold helix (Scott, 1995). As the secondary structure of an HA molecule is flat and tape-like, there is little problem with different HA molecules interacting to form large complexes. As both sides of the HA are identical but anti-parallel, the complex is able to grow from both sides. It has been observed that at low concentrations (< 1 mg/L) of HA, small honeycomb islands of HA form (Scott, 1995). These small networks of HA show no molecular tails or ends and are essentially infinite networks with each molecule connected to every other molecule. As the HA concentration in a solution increases, the network branches become thicker until at concentrations of > 1 mg/ml, as found in synovial fluid, the HA network forms continuous sheets and tubules (Scott, 1995). HA binds to the collagen fibrillar network and determines the retention of aggrecan in the ECM.

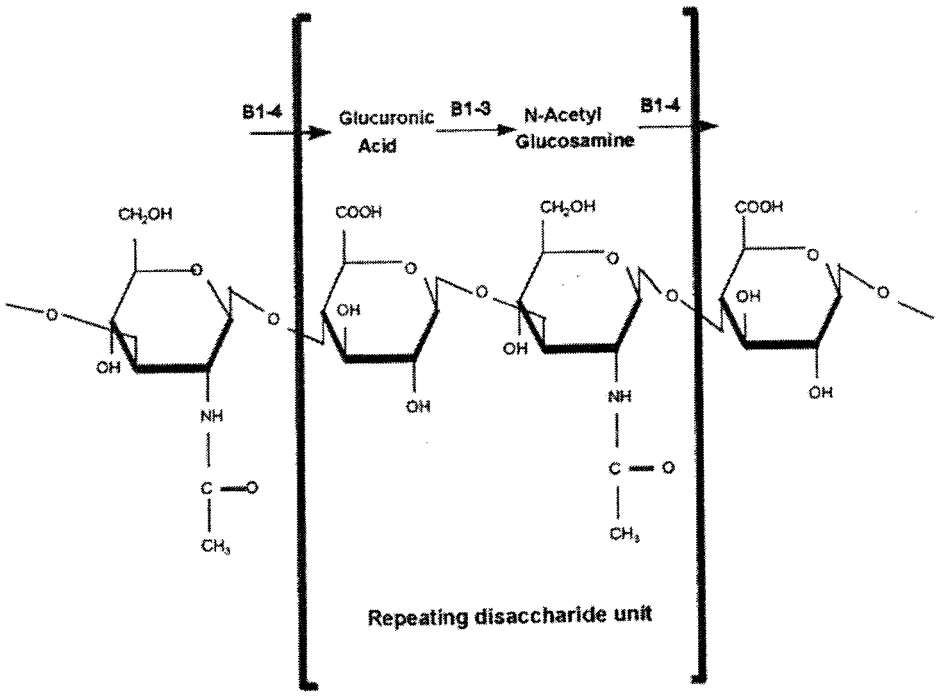


Figure 1.6: Structure of hyaluronan (hyaluronic acid; HA).
HA, a glycosaminoglycan (GAG), is composed of a repeating disaccharide straight chain polymer with the structure (β-1,3-*N*-acetyl-D-glucosamine-β-1,4-D-glucuronic acid)$_n$. Modified from Zubay, 1998.

1.2.5.1.1 Biosynthesis and distribution of hyaluronan

HA is synthesised by *Streptococci* as well as being present in the connective tissue of all animals so far investigated (Knudson *et al.*, 1993). The HA chain is produced by the addition of sugar residues from their UDP derivative onto the reducing end of the chain. This donation of sugar occurs in the plasma membrane by the enzyme HA synthatase (DeAngelis, 1999) and as the chain elongates, the non-reducing end is translocated into the pericellular space. This synthesis differs from that of other GAGs, which are synthesised in the Golgi and the sugars added to the non-reducing end of the chain (Laurent *et al.*, 1992).

HA is found in all ECM structures but is especially prevalent in soft connective tissue and the central nervous system. In the specialised ECM of cartilage tissue, HA forms large aggregates with the protein complex, aggrecan (Knudson *et al.*, 1993). The aggrecan and HA complex is further stabilised by link protein to give a final complex of about 10^8 Da. The aggrecan complex is held within the cartilage by the collagen matrix and is responsible for the visco-elastic properties of cartilage (Knudson and Knudson, 2001).

In many tissues in the body the role of HA is structural, retaining and organising the ECM, as in vitreous humour and Wharton's jelly (umbilical cord). In the joints, the viscous properties of HA are used as a lubricant to prevent wear (Culty *et al.*, 1992; Laurent *et al.*, 1992; Knudson *et al.*, 1993). In cells, it is thought to act as both a protective barrier and a lubricant allowing layers of opposing cells to slide over one another and preventing them sticking together (Heldin *et al.*, 1993; Knudson *et al.*, 1993).

The role of HA in normal cellular functions is well documented (Laurent *et al.*, 1992; Knudson *et al.*, 1993; Lesley *et al.*, 1993). Research has shown HA to be important in cellular proliferation during embryogenesis and increased HA synthesis is required for fibroblast proliferation (Brect *et al.*, 1986). It has been proposed that the increased production of HA at the plasma membrane causes a slight loosening of the cells from the supporting ECM making cell division easier

(Laurent *et al.*, 1992). An increase in the amount of HA contained within the ECM has been correlated with cell migration in embryogenesis, limb regeneration, wound healing and tumour cell invasion. A matrix that is rich in HA has an expanded and highly hydrated structure through which it is easier for cells to migrate (Knudson *et al.*, 1993).

It has been shown that the presence of a HA coat protects cells from cytotoxic lymphocytes and offers some resistance to viral infection. This method of protection is also used by some *Streptococci* infections, to defend against macrophage attack (Laurent *et al.*, 1992).

When structural work was performed on HA it was observed that complexes were formed between phospholipids through the hydrophobic patches on HA. This has been suggested to be important for mopping-up harmful lipidic inflammatory mediators, such as platelet-activating factor, so preventing inflammation spreading too far from the site of damage (Scott, 1995).

It is well documented that there is an increase in the production of HA in cancer and HA interactions have been implicated in many of the different stages of tumour development and metastasis Heldin *et al.*, 1993; Bartolazzi *et al.*, 1994). The ability to bind HA in the ECM could give a tumour an advantage when binding at distant sites. This advantage could take the form of an improved affinity for the initial anchorage site, or it could give the tumour cell access to growth factors that are sequestered by HA and the ECM proteoglycans (Fujita *et al.*, 1994).

1.2.5.2 Heparin / Heparan sulphate

Heparin is a variably sulphated GAG that consists predominately of alternating α-1,4-linked residues of D-iduronate-2-sulphate and *N*-sulpho-D-glucosamine-6-sulphate. It has an average of 2.5 sulphate residues per disaccharide unit, which makes it the most negatively charged polysaccharide in mammalian tissues. Heparin, in contrast to CS, DS, KS and HA is not a constituent of connective tissue, but occurs almost exclusively in the intracellular granules of the mast cells that line arterial walls, especially in the liver, lungs and skin (Fath *et al.*, 1999;

Esko and Lindahl, 2001). Heparin is probably more famous for its anticoagulant effects of blood and is in wide clinical use, but it is known that the anticoagulant properties of the vascular network are also due to the presence of HS on the surface of endothelial cells, since mast cell defficient mice have a normal anticoagulant physiology (Marcum *et al.*, 1986).

HS, a ubiquitous cell surface component as well as an extracellular substance in blood vessel walls and brain, resembles heparin but has a far more variable composition with fewer *N*- and *O*-sulphate groups and more *N*-acetyl groups (Esko and Lindahl, 2001).

HS, as with other GAGs, is an anionic polysaccharide. It consists of many disaccharide units, which are sulphated at varying intervals along the chains. HS is characterised by alternating uronic acid (D-glucuronic acid or L-iduronic acid) and D-glucosamine units (Kjellén and Lindahl, 1991; Esko and Lindahl, 2001). The size of an individual chain is not normally greater than 50 kDa, but some chains can reach a size of 100 kDa.

In addition, several features allow HS to be distinguished. Sulphation occurs at a different stage; in CS/KS biosynthesis, sulphation occurs at the same time as chain elongation, whereas sulphation of HS mainly occurs post-elongation (Lindahl *et al.*, 1989). In addition, most of these sulphates are ester-linked but the amino sulphate modification is unique to HS.

1.2.5.3 Chondroitin sulphate

The sulphated anionic polysaccharide CS is an abundant component of extracellular matrices, especially cartilage. However, it is also found on cell surfaces, in neural tissues and in invertebrates (Lauder *et al.*, 2000). CS chains comprise alternating *N*-acetyl-D-galactosamine and D-glucuronic acid, which may be sulphated on the C4 or C6 position of its *N*-acetyl-D-galactosamine residues to produce chondroitin-4-sulphate and chondroitin-6-sulphate, respectively (Scott and Heatley, 1996). The two CSs occur separately or in mixtures depending on the tissues (Scott and Heatley, 1996).

1.2.5.4 Dermatan sulphate

DS, which is so named because of its prevalence in skin, contains alternating *N*-acetyl-D-galactosamine and D-glucuronic acid residues and differs from chondroitin-4-sulphate only by an inversion of configuration about C5 of the β-D-glucuronate residues to form α-L-iduronate (Scott and Heatley, 1996).

1.2.5.5 Keratan sulphate

KS contains alternating β-1,4-linked D-galactose and *N*-acetyl-D-glucosamine-6-sulphate residues. It is the most heterogeneous of the major GAGs in that its sulphate content is variable and it contains small amounts of fucose, mannose, *N*-acetylglucosamine and sialic acid (Yamagishi *et al.*, 2003).

1.2.5.6 Collagens

Collagen is the most common protein of the vertebrate body and has a unique structure. There are more than 19 different types of collagen and at least 10 additional proteins with collagen-like domains (Prockop and Kivirikko, 1995), forming a wide range of structures. These include fibrils (types I, II, III, V, and XI) that are found in most connective tissues and network-like structures (types VIII, X, and IV). Type IV collagen is a major structural component of basement membranes. Short-chain collagens (types IX, XII, XIV, XVI, and XIX) presenting smaller quantities, attach to the surfaces of pre-existing fibrils. In addition, other types of collagens form beaded filaments (type VI), anchoring fibrils (type VII) and some have transmembrane domains (types XIII and XVII; van der Rest and Garrone, 1991; Prockop and Kivirikko, 1995; Linsenmayer, 1991).

The primary structure of collagen consists of polypeptide chains (α-chains) of about 1000 amino acids, comprising a characteristic tripeptide sequence (glycine-X-Y). Each α-chain is coiled into a left-handed helix. Three α-chains intertwine tightly into a right-handed helix to form a rope-like structure that is stabilised by interchain hydrogen bonds. The triple-helical conformation is wound such that the peptide bonds linking the adjacent amino acids are buried within the interior of the molecule, thus making it highly resistant to attack by general proteases (Prockop and Kivirikko, 1995). Glycine, placed at every third residue of the tripeptide sequence, is small enough to occupy the crowded interior of the helix.

The imino acids proline and 4-hydroxyproline are frequently (about 25 %) in the X and Y positions, respectively. Hydroxyproline is formed as a post-translational modification of proline. They stabilise the polypeptide chains because of the stereochemical restrictions of their imino acid rings. The triple helix is further stabilised by hydrogen bonds and water bridges, many of which require the presence of 4-hydroxyproline (Brodsky and Ramshaw, 1997). The side chains of amino acids in the X and Y positions are on the surface of the molecule. The multiple clusters of hydrophobic and charged side chains direct self-assembly into precisely ordered structures (Prockop and Kivirikko, 1995).

Type II collagen is the single most abundant protein in normal articular cartilage, comprising approximately 90 % of the total collagen. It is a fibrillar collagen, consists of three identical α-chains, and is synthesised as a precursor form, procollagen, which contains propeptides needed for correct fibril assembly. After secretion, the globular propeptide domains are proteolytically removed from the ends of the procollagen, and the collagen molecules assemble into cross-striated fibrils in which each molecule is staggered relative to its nearest neighbour along the axis of the fibril. Covalent cross-links form between the collagen molecules, which increase tensile strength and stability. In normal adult articular cartilage, type II collagen together with smaller amounts of other collagen types (VI, IX, X and XI) form the extracellular framework. Type II, XI and IX collagens that are cartilage specific, are incorporated into the same collagen fibril to form a fibrillar meshwork that gives the tissue its form and tensile strength (Buckwalter and Mankin, 1997).

1.3 Carbohydrate-active enzymes

The molecular recognition of carbohydrates by proteins is of fundamental importance in numerous biological processes, including cell-cell recognition, cellular adhesion, and host-pathogen interactions (Simpson *et al.*, 2000). Carbohydrate-active enzymes are often multi-modular enzymes involved in the synthesis and degradation of glycosides such as di-, oligo- and polysaccharides. Thus, understanding the structural and biochemical basis of these enzymes, along with the ligand specificity of carbohydrate-binding proteins is critical.

1.3.1 Carbohydrate-active enzymes classification

A classification system for the carbohydrate-active enzymes has been introduced (Henrissat and Coutinho, 2000; http://afmb.cnrs-mrs.fr/CAZY/) in families based on amino acid sequence similarities (Henrissat, 1991; Henrissat *et al.*, 1998; Coutinho and Henrissat, 1999). This present classification procedure is based upon hydrophobic cluster analysis (Henrissat, 1991; Henrissat and Bairoch 1996; Coutinho and Henrissat, 1999), a powerful amino acid sequence comparative technique that can detect structural similarities despite low primary amino acid similarities. Glycoside hydrolases (GHs) are classified into highly related 'families' of enzymes, the members of which share primary sequence homology and hence the same overall 3D fold. This classification system for GHs has many advantages compared to the traditional EC classification system for GHs (EC 3.2.1.x), which assigns enzymes primarily on substrate specificity (Coutinho and Henrissat, 1999). A significant advantage of classification according to sequence similarities is that it allows logical grouping of enzymes of different EC numbers into polyspecific families and offers insights into the divergent evolution of enzyme families. Conversley, some enzymes that can be grouped by function have been shown to belong to several distinct families and thus reflect convergent evolution (Henrissat, 1991; Davies and Henrissat, 1995; Tomme *et al.*, 1995a; Campbell *et al.*, 1997, Henrissat and Davies, 1997; Coutinho and Henrissat, 1999). The CAZy server describes the modular structure of carbohydrate-active enzymes and has been created to provide access to updated classifications of glycosidases and transglycosidases, which are collectively known as the GHs, glycosyltransferases (GTs), polysaccharide lyases (PLs), and carbohydrate esterases (CEs), and their associated modules (carbohydrate-binding modules (CBMs)) in families. Underlying this classification was the prediction that sequence determines structure, and thus useful structural, evolutionary and even mechanistic information can be derived from amino acid sequence alone (Campbell *et al.*, 1997; Henrissat and Davies, 1997; Davies, 1998). Thus, all members of a family share common properties that can be predicted if the property is determined for one or more of the family members. One of the striking examples is the stereochemistry of hydrolysis; a family can contain enzymes with endo- and exo- modes of action, and even different substrate specificities, but the

catalytic machinery and thus the stereochemistry of hydrolysis remains conserved e.g. all family 5 members are retaining enzymes, whereas all family 6 members are inverting enzymes (Warren, 1996). The proteins in this thesis belong to GH69 class and CBM6 class of the carbohydrate-active enzymes.

1.3.1.1 Glycoside hydrolases

The GHs are a class of enzymes that are responsible for the hydrolysis of glycosidic bonds. Such glycosidic linkages occur in a wide range of contexts, including polysaccharides, oligosaccharides, glycolipids, glycoproteins, lipopolysaccharides, proteoglycans, saponins, and a range of other glycoconjugates (Vocadlo & Withers, 2000). They perform innumerable biological roles, including the breakdown of plant and animal polysaccharides into units small enough to be transported into the host organism and utilised as a source of energy and carbon. Corresponding to this diverse collection of substrates there is a very large assortment of GHs responsible for their selective hydrolysis. Currently, (May 2004) there are 95 families of GHs and a large amount of effort has been expended on structural studies of these enzymes, with the result that 3D X-ray crystal structures are now available for representatives of at least 55 of these families (Henrissat and Coutinho, 2000).

1.3.1.1.1 Molecular architecture of GHs

A common feature of GHs is that they often exhibit a modular structure, consisting of discrete domains that are linked by sequences rich in hydroxy amino acids (Gilbert *et al.*, 1990). As well as their multiplicity, analysis of the molecular architecture of hydrolases has revealed a further level of complexity. It has been shown that the enzymes themselves are often modular in structure, comprising of a catalytic module joined by one or more ancillary modules, often via a flexible linker sequence (Tomme *et al.*, 1995a). As the linker regions are very sensitive to proteolysis, it was discovered that the individual modules were functionally independent (van Tilbeurgh *et al.*, 1986; Tomme *et al.*, 1988; Gilkes *et al.*, 1988). CBMs have since been found to be the most common type of ancillary module present in hydrolases (Tomme *et al.*, 1995a). Recombinant DNA technology has been used to generate, as discrete entities, the individual modules. The identity of modules in newly sequenced genes can now, usually, be determined simply by

comparison of their primary structures with those of modules whose function is already known.

1.3.1.1.2 Active site topologies of GHs

The 3D structure data for the catalytic modules of GHs have revealed three general active site topologies: pocket, cleft (or groove) and tunnel. These topologies exist in disaccharidases, endo- and exo-enzymes respectively (Figure 1.7; Davies and Henrissat, 1995). In general, the pocket topology is optimal for the recognition of glycoside chain ends and is typical of disaccharidases such as glucosidases and xylosidases. The cleft or groove topology, with an 'open' active site, is typical of endo-acting enzymes such as endo-glucanases and xylanases. The cellulose and xylan chains, respectively, bind randomly within the cleft, in agreement with the random internal cleavage of the substrate and the generation of oligosaccharides of variable length. The tunnel topology, found in exo-acting cellobiohydrolases, is similar to the cleft or groove topology, but with additional polypeptide loops that cover the top of the active site forming tunnel-like structures. The active site tunnel of exo-acting enzymes accommodates a single glycan chain (Davies and Henrissat, 1995).

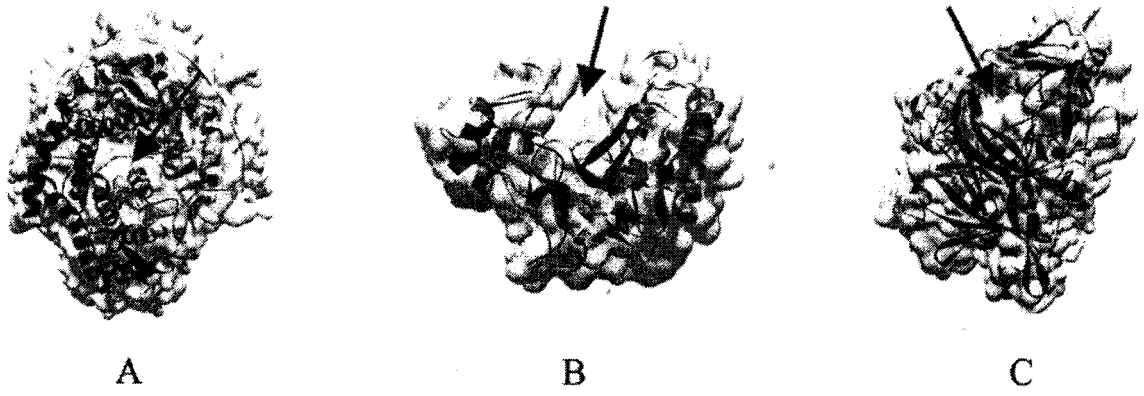


Figure 1.7: Active site topologies of glycosyl hydrolases.
The pocket (A: glucoamylase from *Aspergillus awamori*; Aleshin *et al.*, 1992); cleft (B: endo-1,4-β-D-glucanase from *Thermomonospora fusca*; Spezio *et al.*, 1993); and tunnel (C: cellobiohydrolase from *Trichoderma reesei*; Divne *et al.*, 1994). Active site (indicated by arrows) topologies displayed. Taken from Davies and Henrissat (1995).

1.3.1.1.3 Clans

The determination of the 3D structures for representatives of more than 55 GH families has led to the observation that some families display very similar overall structures and can be grouped into clans (Henrissat and Davies, 1997; Table 1.2) or superfamilies (Jenkins *et al.*, 1995). The structural conservation within a clan suggests that all members have evolved from a single progenitor sequence and although the primary structures of the enzymes have diverged significantly, the regions that are subject to the most intense conservation pressure, i.e. those carrying the catalytic residues, will have remained conserved (Henrissat *et al.*, 1995; Coutinho and Henrissat, 1999). In the absence of mechanistic data, enzyme families within a given clan can be inferred as having the same catalytic residues and also the same catalytic mechanism, i.e. retaining or inverting, when the information exists for other clan members (Warren, 1996; Henrissat, 1998; Coutinho and Henrissat, 1999).

Clan	GH Families
GH-A	1,2,5,10,17,26,30,35,39,42,50,51,53,59,72,79,86
GH-B	7, 16
GH-C	11, 12
GH-D	27, 36
GH-E	33, 34, 83
GH-F	43, 62
GH-G	37, 63
GH-H	13, 70, 73, 77
GH-I	24, 46, 80
GH-J	32, 68
GH-K	18, 20
GH-L	15, 65
GH-M	8, 48
GH-N	28, 49

Table 1.2: Clans of glycoside hydrolase families.

1.3.1.1.4 Overview of the mode of action of GHs

The efficient and complete degradation of polysaccharides, whether they are from plant or animal cells, requires organisms to produce an extensive array of endo- or exo-acting enzymes. GHs are described as being either endo- or exo- in their mode of action, depending on whether the substrate is attacked at random internally, or from one of the ends, respectively. GHs release distinct products due to their different modes of action on their target substrates (Warren, 1996). Endo-acting hydrolases randomly cleave glycosidic bonds within polysaccharide chains. In contrast, exo-acting hydrolases remove one or more residues from either the reducing or non-reducing termini of oligosaccharide. Thus, *in vivo*, both endo- and exo-acting enzymes would act in synergy, with the products of endo-acting enzymes providing substrates for exo-acting enzymes, leading to more efficient degradation.

1.3.1.1.5 Catalytic mechanisms

The majority of enzymes comprising the 95 families of GHs so far defined catalyse the hydrolysis of glycosidic bonds with either retention or inversion of anomeric configuration at C1 of the bond being broken, via general acid/base assisted catalysis (Sinnott, 1990; Warren, 1996). Figure 1.8 illustrates these two catalytic mechanisms. Both mechanisms require two key catalytic carboxylic acid residues in the active site of the enzyme; however, the roles of these two residues vary depending on the mechanism. In the vast majority of GHs that have been studied, only aspartic acid and/or glutamic acid residues comprise the catalytic amino acids.

1.3.1.1.5.1 Inverting mechanism

Inverting enzymes catalyse hydrolysis by a single displacement mechanism resulting in inversion of configuration of the anomeric carbon (Davies and Henrissat, 1995). In inverting hydrolases, one of the key carboxylic acid residues acts as a general acid and the other as a general base. The mechanism of inverting enzymes (Figure 1.8 (a)) involves a single step in which the general acid catalyst protonates the oxygen of the glycosidic bond, while the general base simultaneously promotes the formation of an hydroxyl ion from a water molecule,

which performs a nucleophilic attack on the anomeric carbon of the glycoside. As a result, the glycosidic bond is cleaved and a product with the opposite stereochemistry to the substrate is generated. The two catalytic residues of inverting enzymes are normally separated by ~ 10 Å to allow the accommodation of a water molecule between the general base and sugar (Davies and Henrissat, 1995; Davies, 1998).

1.3.1.1.5.2 Retaining mechanism

Retaining enzymes hydrolyse glycosidic bonds via a double-displacement mechanism, leading to a net retention of anomeric configuration of C1 (Davies and Henrissat, 1995). In retaining hydrolases, one of the two key carboxylic acid residues functions as a general acid and a general base (acid/base catalyst) while the other acts as a nucleophile (McCarter and Withers, 1994). The mechanism for retaining enzymes involves a two-step process (Figure 1.8 (b)). In the first step, the acid/base residue acts as an acid and protonates the oxygen of the glycosidic bond, resulting in bond cleavage, and the nucleophile residue attacks the anomeric carbon of the glycosidic bond to form a covalent glycosyl-enzyme intermediate (Tull and Withers, 1994; Davies, 1998). The acid/base catalyst then functions as a general base and removes a proton from a water molecule, which attacks the glycosyl-enzyme intermediate, resulting in the generation of a sugar product with the same configuration as the substrate, and the release of free enzyme. It has been demonstrated that the distance between the two catalytic residues of retaining enzymes is ~ 5.5 Å, which positions the nucleophilic catalytic base in close proximity to the sugars anormeric carbon to enable the formation of the glycosyl-enzyme intermediate (Tull and Withers, 1994; Davies and Henrissat, 1995; Davies, 1998).

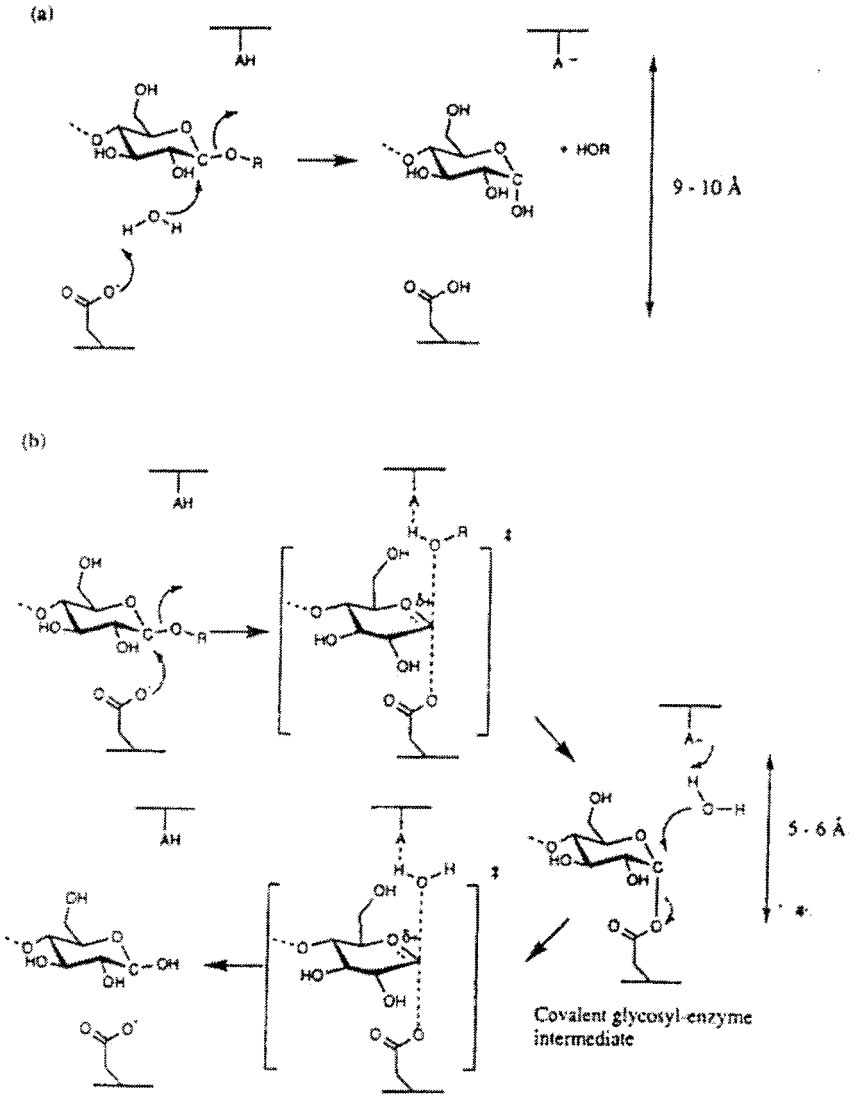


Figure 1.8: Inverting and retaining catalytic mechanisms of GHs.
The 'inverting' (a) and 'retaining' (b) reaction mechanisms of GHs. Taken from Davies (1998).

1.3.1.1.6 Plant cell wall hydrolases

A wide variety of bacteria and fungi synthesise numerous enzymes that hydrolyse PCWs due to the complexity of the substrate. These PCW-degrading microorganisms can be found in any environment where the substrate accumulates, whether it is under aerobic or anaerobic conditions. A consortium of enzymes, which include cellulases, hemicellulases, such as xylanases, and pectinases, mediate this process of degradation.

Some organisms can exploit the plant biomass as a source of carbon and energy, as they produce cellulase systems that can efficiently digest cellulose in the PCWs, exemplified by *Clostridia* sp. (Hazlewood and Gilbert, 1993). In contrast, other microorganisms possess an enzyme system that has relatively limited hydrolysis capacity, example of which includes the plant bacterial pathogen *Erwinia chrysanthemi* (Tomme *et al.*, 1995c).

It is still unclear why a single cellulolytic organism produces a wide range of PCW hydrolases, many of which have overlapping substrate specificities. For example, seven different cellulases, three xylanases and seven other related enzymes have been studied in *Cellvibrio japonicus* (formerly *Pseudomonas cellulosa*; Hazlewood and Gilbert, 1993; Nagy *et al.*, 2000). The bacterium *Clostridium thermocellum* produces 15 cellulases and five xylanases (Béguin and Aubert, 1994). The bacterium *Cellulomonas fimi* produces six cellulases and four xylanases (Clarke *et al.*, 1991; Warren, 1996). The explanation for the multiplicity of hydrolases is likely to lie in the structural complexity of the PCW. The rationale for the acquisition of such large numbers of a similar type of enzyme by a single organism is at present unclear. In the case of xylanases, the multiplicity of enzymes is probably due to the heterogeneity of the substrate, a wide variety of xylanases are required, each with different specificities in relation to the presence and composition of side chains, in order to achieve total hydrolysis of the various types of xylan present in the environment.

Although some PCW hydrolases comprise a catalytic module only, the majority are multi-module enzymes with the number of modules in a single enzyme ranging from two to six or more, one of which is always a catalytic module (Warren, 1996). The various domains, however, are not located in equivalent positions in the different enzymes. For example, unrelated catalytic modules can be found linked both N- and C-terminally to identical types of CBMs. Thus, it is apparent that extensive domain shuffling has occurred during the course of evolution of PCW hydralases (Gilbert and Hazlewood, 1993).

1.3.1.1.6.1 Cellulases

Cellulases, which are involved in the total hydrolysis of the β-1,4-glycosidic bonds in cellulose, belong to three groups and they exhibit different substrate specificity and mode of action, resulting in distinct products (Warren, 1996). Since cellulose cannot enter microbial cells, cellulolytic enzymes are either secreted into the environment or attached to the outer surface of the cells themselves. Endo-β-1,4-glucanases catalyse the hydrolysis of internal glycosidic bonds in cellulose chains at random. Exo-β-1,4-glucanases (also called cellobiohydrolases) sequentially release cellobiose either from the reducing or non-reducing ends of cellulose chains and β-glucosidases hydrolyse cellobiose and some cellooligosaccharides to glucose (Coughlan, 1985; Gilbert and Hazlewood, 1993; Tomme *et al.*, 1995c).

1.3.1.1.6.2 Xylanases

Microorganisms that produce cellulases also synthesise enzymes capable of hydrolysing hemicelluloses, especially xylan, probably because of the close proximity of hemicellulose to cellulose in the PCW. As noted earlier, hemicellulose usually has a more accessible structure than crystalline cellulose, with the result that most hemicellulases have specific activities ~ 2-3 fold greater than cellulases (Gilbert and Hazlewood, 1993). Due to the heterogeneity of most hemicelluloses, their efficient hydrolysis requires a repertoire of enzymes that act co-operatively. In the case of xylan, these enzymes include β-1,4-endoxylanases, which cleave β-1,4-glycosidic bonds within the xylan backbone; β-xylosidases, which hydrolyse xylobiose to xylose; α-arabinofuranosidases, which cleave arabinose side chains; acetylxylan esterases, which release acetate groups; and

finally α-glucuronidases, which remove glucuronic acid side chains from the xylan backbone (Sunna and Antranikian, 1997; Gilbert and Hazlewood, 1993; Figure 1.9). All of these enzymes are known to act in synergy to degrade what is often a highly complex polysaccharide containing a large number of side chain substitutions. For example, many endoxylanases are unable to cleave bonds in the xylan backbone if the xylose residues are substituted. Thus, the actions of the side chain cleaving enzymes are required before total hydrolysis can be achieved. In turn, the shorter substituted oligosaccharides released by the endoxylanases are the preferred substrates for many side chain cleaving enzymes (Sunna and Antranikian, 1997).

Similar to cellulolytic organisms, xylanolytic organisms also produce a range of xylan-degrading enzymes. For instance, *C. japonicus* synthesises four xylanases (Xyn10A, Xyn10B, Xyn10C and Xyn11A) (Millward-Saddler *et al.*, 1995). In addition, this organism was found to produce two ancillary enzymes, an arabinofuranosidase (Abf62A) and an acetyl esterase (XynD). *C. thermocellum* produces five xylanases (Xyn10B, Xyn10C, Xyn10D, Xyn11A, Xyn11B) (Fontes *et al.*, 1995; Fernandes *et al.*, 1999).

1.3.1.1.6.3 Other hemicellulases

To utilise the PCW efficiently, microorganisms also synthesise other hemicellulases, which requires the action of a consortia of main chain, side chain and deacetylating enzymes similar to those responsible for xylan hydrolysis, such as mannanases, mannosidases, arabinanases and arabinofuranosidases. For example, the β-1,4-linked mannose backbone of mannan is cleaved randomly by endo-β-1,4-mannanases to produce mannobiose and longer mannooligosaccharides, from which mannose is removed by β-1,4-mannosidases (Hazlewood and Gilbert, 1998). The complete degradtion of glucomannans requires the action of glucosidases, mannanases and mannosidases (Hazlewood and Gilbert, 1998).

1.3.1.1.6.4 Pectinases

Pectinases are a group of enzymes including hydrolases and lyases, which degrade the pectin component of PCWs (Brown *et al.*, 2001; Charnock *et al.*, 2002a).

Rhamnogalacturonan hydrolases and rhamnogalacturonan lyases degrade the rhamnogalacturonan regions of pectin (McKie *et al.*, 2001), and exo- and endo-polygalacturonases, pectin lyases and pectate lyases catalyse the degradtion of polygalacturonan regions of the acidic polysaccharide (Brown *et al.*, 2001; Charnock *et al.*, 2002a).

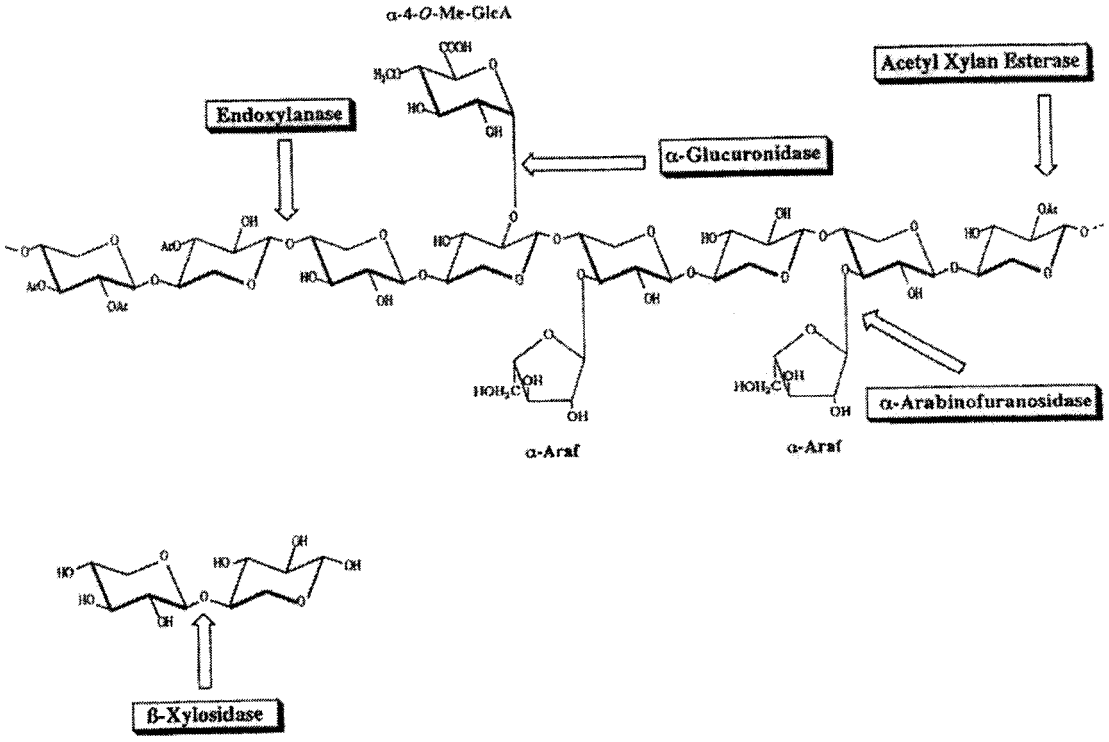


Figure 1.9: Structure of a hypothetical xylan chain showing the major types of possible substituents and the enzymes that cleave them.
Abbreviations are: Ac, acetyl; Araf, arabinofuranose; MeGlcA, methylglucuronic acid. Taken from Sunna and Antranikian (1997).

1.3.1.1.7 Glycosaminoglycan hydrolases

A plethora of bacteria, bacteriophage, fungi, invertebrates and vertebrates synthesise numerous enzymes that hydrolyse GAGs due to the wide distribution in nature of the substrates. These GAG-degrading organisms can be found in any source where the substrates are found. A variety of enzymes, which include hyaluronidases and heparanases, mediate this process (Table 1.3). The degradation of HA and other GAGs by hydrolases and lyases may facilitate the invasion of the tissues of an animal host by bacteria (Hynes and Walton, 2000). Possession of HA-degrading activities may also facilitate bacterial adhesion and colonisation (Polissi *et al.*, 1998). Degradation of GAGs occurs via two possible mechanisms: hydrolysis and lytic β-elimination, with the exception of keratan sulphate, which can only be degraded hydrolytically (Huang *et al.*, 2001).

1.3.1.1.7.1 Hyaluronidases

The term hyaluronidase was introduced to denote enzymes that degrade HA (Frost *et al.*, 1996; Csóka *et al.*, 1997; Menzel and Farr, 1998). However, the term HA-degrading enzymes would be more appropriate as enzymes that degrade HA are comprised of hyaluronidases (i.e. GHs) and hyaluronate-lyases (i.e. PLs). For the purposes of this thesis, hyaluronidases will refer to hydrolases and not the hyaluronate lyases. Hyaluronidases were first identified as a 'spreading factor' in an extract from mammalian testes that facilitated diffusion of subcutaneously injected dyes and drugs (Frost *et al.*, 1996). It became apparent thereafter that the spreading factor was an enzyme and that its substrate was HA (Chain and Duthie, 1939; 1940; Frost *et al.*, 1996; Csóka *et al.*, 1997). However, prior to this, an enzyme that degraded HA had been identified in bacteria (Meyer *et al.*, 1937; Frost *et al.*, 1996). Hyaluronidase is produced by groups A, B, C and G streptococci, staphylococci, candida and clostridia, plus many others (Prescott *et al.*, 1996; Hynes and Walton, 2000). This enzyme hydrolyses HA; a constituent of the intercellular ground substance that cements cells together, allowing for penetration and spread of pathogens.

Different types of hyaluronidases hydrolyse the long chains of alternating units of *N*-acetylglucosamine and D-glucuronic acid present in HA, the most abundant mucopolysaccharide of vertebrate connective tissue (Gmachl and Kreil, 1993). These enzymes are widely distributed in nature; they have been isolated from such diverse sources as mammalian testes, salivary glands of leeches, venoms of snakes, bees and wasps, and different bacteria (Meyer, 1971). A wide variety of bacteria produce enzymes capable of degrading HA. Enzymes produced by the Gram-negative organisms are periplasmic rather than being excreted to the extracellular environment, and so are less likely to play a role in pathogenesis, in contrast to the Gram-positive organisms (Hynes and Walton, 2000).

1.3.1.1.7.1.1 Bacteriophage-associated hyaluronidases

Bacteriophage isolated from a strain of *S. pyogenes* possesses a hyaluronidase unrelated to the bacterial enzyme (GH69) (Hynes and Ferretti, 1989; Hynes *et al.*, 1995). *Streptococcus equi* has also been shown to encode a bacteriophage-associated hyaluronidase (Timoney *et al.*, 1982). It has been suggested that the biological function of bacteriophage hyaluronidase is to degrade the HA capsule of the bacterial host cell, thereby facilitating attachment of phage to the cell wall (Hynes *et al.*, 1995, Hynes and Walton, 2000).

1.3.1.1.7.1.2 Fungal hyaluronidases

Infections caused by candida are the most common fungal infections affecting humans and various species of candida produce hyaluronidases (Shimizu *et al.*, 1995). The enzymes have currently not been characterised, nor has it been established whether these enzymes are virulence factors or whether they play a role in pathogenesis (Shimizu *et al.*, 1996).

GH Family & clan	Members	Representatives	3D structure available for family	Known activity	Known mechanism	References
16 (GH-B)	331	*Flavobacterium keratolyticus*	Yes	Keratan sulphate-endo-1,4-β-galactosidase	Retaining	Leng *et al.*, 1998; Yamagishi *et al.*, 2003
56	42	*Apis mellifero* (Honey bee venom) *Homo sapiens* (Sperm protein PH-20) *Xenopus laevis* (XKH1 kidney) *Synanceja horrida* (HYA1 Stonefish)	Yes	Hyaluronidase	Retaining	Gmachl & Kreil, 1993 Lin *et al.*, 1993; Gmachl *et al.*, 1993 Reitinger *et al.*, 2001 Sugahara *et al.*, 1992
69	21	*Streptococcus pyogenes* (Bacteriophage H4489A) *S. pyogenes* SF370 (ORFs HylP1, HylP2, HylP3)	No	Hyaluronidase	Unknown	Hynes & Ferretti, 1989 Ferretti *et al.*, 2001
79 (GH-A)	19	*Homo sapiens* *Bos taurus* (Bovine)	No	Heparanase / endo-β-glucuronidase	Retaining (Inferred)	Dempsey *et al.*, 2000 Kizaki *et al.*, 2001
84	22	*Clostridium perfringens* (NagH) *S. pyogenes* SF370	No	Putative hyaluronidase / *N*-acetyl-β-glucosaminidase	Unknown	Shimizu *et al.*, 2002 Ferretti *et al.*, 2001
88	28	*S. pyogenes* SF370 (Ugl)	Crystallisation note reported	Δ α-4,5-unsaturated glucuronyl hydrolase	Unknown	Hashimoto *et al.*, 1999

Table 1.3: GH families associated with the hydrolysis of glycosaminoglycans.

Based on data derived from CAZy (http://afmb.cnrs-mrs.fr/CAZY/).

1.3.1.1.7.1.3 Invertebrate hyaluronidases

The venom of bees, wasps and hornets contain a hyaluronidase that is used as a spreading factor facilitating the diffusion of other venom components (Kemeny *et al.*, 1984; Gmachl and Kreil, 1993; Lu *et al.*, 1995). A hyaluronidase has been cloned, expressed and purified from the venom of the honey bee, *Apis mellifera* (GH56) (Kemeny *et al.*, 1984; Gmachl and Kreil, 1993). Interestingly, the hyaluronidase from bee venom glands exhibited significant homology to PH-20, a membrane protein of sperm (GH56)(Gmachl and Kreil, 1993; Section 1.3.1.1.7.1.5).

The hyaluronidase from the leech (*Hirudo medicinalis*) is specific for HA and has no reactivity on CS (Yuki and Fishman, 1963). It is evident that the leech hyaluronidase has a potential role as a spreading factor (Frost *et al.*, 1996), facilitating the spread of other agents that help the leech obtain blood (Csóka *et al.*, 1998).

Infective hookworm (*Anclystoma caninum*) larvae are known to secrete a hyaluronidase during penetration of the skin (Hotez *et al.*, 1992; 1994). A hyaluronidase has also been detected in crustaceans such as the Antarctic krill (*Euphausia superba*; Karlstrom and Ljunglof, 1991).

1.3.1.1.7.1.4 Vertebrate hyaluronidases

Hyaluronidases are also present in venoms of snakes, certain lizards and a species of fish. Venom from the stonefish *Synanceja horrida* contains a hyaluronidase that has been isolated and characterised (Sugahara *et al.*, 1992). As in the case of the leech, hyaluronidase from the stonefish may assist in the rapid spread of venom through tissues (Frost *et al.*, 1996).

Hyaluronidases are present in virtually all snake venoms, but only a few have been characterised (Girish *et al.*, 2002). Venom hyaluronidases have been shown to act as spreading factors and by themselves are non toxic. Degradation of HA in the ECM of local tissues is thought to be the key event in the enzyme mediated spreading process during snakes' envenomation (Girish *et al.*, 2002).

In mammals, hyaluronidases have been found in a range of sites. A protein termed PH-20 (GH56), present on the plasma membrane and in the matrix of the sperm acrosome, was shown to be the sperm hyaluronidase and that it had homology with bee venom hyaluronidase (Gmachl *et al.*, 1993). The cumulus cells are embedded in a matrix rich in HA that is formed prior to ovulation, and the sperm hyaluronidse PH-20 assists in this penetration process (Lin *et al.*, 1994a; 1994b).

Kidney and liver are the two major organs for HA degradation and hyaluronidases have been isolated from these sources (GH56) (Reitinger *et al.*, 2001 and Joy *et al.*, 1985, respectively). Skin contains more than half the HA of the vertebrate body (Frost *et al.*, 1996) and a hyaluronidase derived from the skin has been identified (Cashman *et al.*, 1969). Hyaluronidase from human placenta has been characterised (Yamada *et al.*, 1977), as well as from vitreous (Schwartz *et al.*, 1996), brain (Polansky *et al.*, 1974), and salivary glands (Tan and Browness, 1968a; 1968b). A hyaluronidase in malignant prostate tissue has been observed that has levels of activity that correlate with cancer progression (Lokeshwar *et al.*, 1996; Frost *et al.*, 1996).

1.3.1.1.7.2 Heparanases

Heparanases are endo-β-D-glucuronidases that degrade HS (GH79) (Ernst *et al.*, 1995; Kizaki *et al.*, 2001). Heparanase activity has been reported in various tumour cells, platelets, and placenta (Nakajima *et al.*, 1983; Freeman and Parish, 1998; Vlodavsky *et al.*, 1999; Kizaki *et al.*, 2001). Studies have shown that heparanase-mediated degradation of HS in the ECM may be critical during cell invasion associated with tumour metastasis and inflammation (Bartlett *et al.*, 1995; Vlodavsky *et al.*, 1999). Heparanse has been shown to be strongly expressed in human placenta (Kussie *et al.*, 1999) and bovine placenta (Kizaki *et al.*, 2001), but the role played during implantation and placental development is unknown.

1.3.1.1.7.3 Keratinases

KS does not have any uronic acid therefore the lyase reaction is not possible, and hydrolases are the only enzymes depolymerising this GAG (Ernst *et al.*, 1995).

KS is degraded by various enzymes including endo-β-galactosidase, keratanase and endo-β-*N*-acetylglucosaminidase (GH16; Yamagishi *et al.*, 2003). Studies have revealed various functions of KS chains in embryogenesis, cancer invasion, immune responses and cartilage metabolism (Ernst *et al.*, 1995; Yamagishi *et al.*, 2003).

1.3.1.1.8 Polysaccharide lyases

There are two classes of enzyme that perform cleavage of the glycosidic bond, the GHs (Section 1.3.1.1) and the PLs. PLs, however, cleave the glycosidic bond via a β-elimination mechanism resulting in the formation of a double bond at the newly formed non-reducing end, which is in contrast to the GHs that cleave the glycosidic bond via a hydrolytic mechanism. Fifteen sequence-based families of PLs have been identified to date, compared to 95 families of GHs. The higher prevalence of the GHs has led to them being extensively studied with a wealth of structural and biochemical data available. This success has not been achieved to the same extent in PLs and is presumably as a consequence of limited substrate specificity. PLs, unlike GHs, require free or methyl esterified C5 substrate carboxylate groups for activity and are thus not as ubiquitous in nature (Sutherland, 1995).

1.3.1.1.8.1 Glycosaminoglycan lyases

A number of bacterial species synthesise GAG lyases, enzymes used to degrade and utilise GAGs as a source of carbon in the bacterium's natural environment (Ernst *et al.*, 1995; Sutherland, 1995).

1.3.1.1.8.1.1 Hyaluronate lyases

Hyaluronate lyases (HA lyases) are produced by many different genera of bacteria, such as *S. pyogenes* (Baker *et al.*, 2002), *Streptococcus agalactiae* (Pritchard *et al.*, 1994; Lin *et al.*, 1994a; 1994b; Baker *et al.*, 1997), *Streptococcus pneumoniae* (Berry *et al.*, 1994) and *Streptomyces hyalurolyticus* (Price *et al.*, 1997) and belong to PL family 8. Bacterial HA lyases have been shown to cleave *N*-acetylglucosaminidic bonds of HA by an elimination mechanism. The products are unsaturated oligosaccharides with a Δ-4,5-uronic acid residue at the non-reducing termini (Baker *et al.*, 2002). Some HA lyases also have a limited ability

to cleave CS and DS in a similar manner, but to a far more limited extent (Baker *et al.*, 1997; Csóka *et al.*, 1998).

Many pathogenic streptococci produce extracellular HA lyases, which are thought to aid the spread of the organism in host tissues (Baker *et al.*, 2002). HA lyase in *Clostridium perfringens* is a putative virulence factor, acting in conjunction with collagenases on connective tissues in the pathogenesis of gas gangrene (Canard *et al.*, 1994). The HA lyase associated with the syphilis bacterium *Treponema pallidum* is known to facilitate invasion and dissemination (Fitzgerald and Gannon, 1983). Interestingly, non-pathogenic strains of treponema do not possess HA-lyase activity (Csóka *et al.*, 1998).

1.3.1.1.8.1.2 Heparin / heparin-sulphate lyases

Two types of enzymes that degrade heparin and HS are the PLs and the GHs (Section 1.3.1.1). Prokaryote degradation of heparin and HS has primarily been studied using enzymes derived from *Flavobacterium heparinium* (Lohse and Linhardt, 1992; Godavarti and Sasisekharan, 1996; Godavarti *et al.*, 1996; Su *et al.*, 1996; Shriver *et al.*, 1998; Nader *et al.*, 1999), which belong to PL family 8, 12 and 13. This bacterial degradation is achieved by the action of three PLs that are capable of acting on linkages present in heparin and HS: heparin lyase I (heparinase I), heparin lyase II (heparitinase II) and heparin lyase III (heparitinase III) (Lohse and Linhardt, 1992). Heparin lyases from *F. heparinium* convert heparin and/or HS to disaccharides through eliminative cleavage of the glucosamine-uronate linkage (Su *et al.*, 1996; Godavarti and Sasisekharan, 1996).

1.3.1.1.8.1.3 Chondroitin lyases

Chondroitin lyases are synthesised by a number of bacterial species that degrade GAGs, namely chondroitin, and belong to PL family 8 as they share significant sequence homology with a number of bacterial hyaluronate lyases. For example, chondroitin lyases include the chondroitin AC lyase and B lyase from *F. heparinium* (Tkalec *et al.*, 2000; Huang *et al.*, 2001), two distinct chondroitin sulphate ABC lyases I and II from *Proteus vulgaris* (Hamai *et al.*, 1997; Huang *et al.*, 2003), chondroitin lyases I and II from *Bacteroides thetaiotaomicron* (Linn *et al.*, 1983) and the chondroitin AC lyase from *Arthrobacter aurescens* (Lunin *et*

al., 2004). Chondroitin lyases cleave the β-1,4-galactosaminidic bonds between *N*-acetylgalactosamine and either D-glucuronic acid or L-iduronic acid and are therefore capable of degrading a variety of GAGs, specifically chondroitin-4-sulphate and chondroitin-6-sulphate, but also DS and HA (Tkalec *et al.*, 2000).

1.3.1.2 Carbohydrate-binding modules

The vast majority of modular PCW hydrolases, such as cellulases and hemicellulases, often comprise of a catalytic module joined to non-catalytic modules by linker sequences. The most common non-catalytic modules associated with PCW degrading enzymes are CBMs (Warren, 1996). Similar to the GHs, CBMs are grouped into 39 families to date (Henrissat and Coutinho, 2000; CAZy server). Interestingly, unlike the GHs, CBMs appear to form families with all members originating from enzymes of either bacterial or fungal origin. Family one contains only fungal and algal CBMs, family 29 contains a fungal CBM, whereas all other families contain exclusively bacterial CBMs (CAZy server). The rationale for this strict separation is unclear, as it is known that lateral transfer of DNA encoding catalytic domains is relatively common between bacteria and fungi (Tomme *et al.*, 1995b).

The majority of functional CBMs show affinity for cellulose, which were previously designated as cellulose-binding domains (CBDs). CBDs exist not only in fungal and bacterial cellulases, but also in xylanases, an arabinofuranosidase, a pectate lyase and two acetyl esterases (Kellet *et al.*, 1990; Ferreira *et al.*, 1993; Brown *et al.*, 2001; Linder and Teeri, 1997). Some CBMs bind xylan and were thus previously designated as xylan-binding domains (XBDs; Black *et al.*, 1995; Dupont *et al.*, 1998; Simpson *et al.*, 1999; Kuno *et al.*, 2000). Some CBMs, although preferentially interacting with xylan, show broad ligand-specificity, with affinity for both amorphous cellulose as well as xylan (Sun *et al.*, 1998; Fernandes *et al.*, 1999; Abou Hachen *et al.*, 2000). All CBMs that bind to xylan have so far been found in xylan degrading enzymes but not cellulases, including several xylanases from different origins, an acetyl xylan esterase and an arabinofuranosidase (Dupont *et al.*, 1998; Simpson *et al.*, 1999; Kuno *et al.*, 2000; Sunna *et al.*, 2000).

Distinct families contain variable numbers of CBMs. The CBM families and representatives of each family involved in cellulose and/or hemicellulose binding are given in Table 1.4. CBM6, comprising 49 members, are found in bacterial GHs, including xylanases, cellulases, α-1,6-mannanases, acetylxylan esterases, cellodextrinases, β-1,3-glucanases and α-agarases. In addition, it should be noted that sequence alignment has revealed the presence of subfamilies in families 2, 3, 4, and 22.

Discrete CBMs are also found in enzymes that act on other insoluble polysaccharides, such as starch and chitin, suggesting that their presence offers significant advantages in the degradation of these insoluble substrates (Coutinho and Reilly, 1993; Blaak and Schrempf, 1995).

1.3.1.2.1.1 Other non-catalytic modules

Although CBMs are by far the most common ancillary modules of GHs, additional non-catalytic modules that include dockerin domains, S-layer-like (SLH) domains and fibronectin type III (Fn3) domains are also present in some GHs. It has been proposed that the SLH domains anchor bacterial hydrolases to the microbial cell surface (Fujino *et al.*, 1993; Lemaire *et al.*, 1998). Fn3 domains may fulfil a linker function, as they are generally located between a catalytic module and a CBM (Hansen, 1992; Little *et al.*, 1994). Alternatively, these domains may be involved in protein-protein or protein-cell interactions within PCW polysaccharide degrading systems (Tomme *et al.*, 1995a). Dockerin domains of glycanases from *C. thermocellum* are involved in the assembly of the cellulosome (Carvalho *et al.*, 2003).

1.3.1.2.1.2 Linker sequences

Individual domains in modular GHs are generally joined by linker sequences that are rich in hydroxyl amino acid residues (Gilkes *et al.*, 1991). These linkers generally consist of repeats of either proline and threonine, serine or glycine residues and appear to have a relatively extended and flexible structure (Srisodsuk *et al.*, 1993; Tomme *et al.*, 1995a). Furthermore, they are often glycosylated to protect against proteolysis (Coggins, 1991).

The precise role of linker sequences is unclear, however, their widespread occurrence indicates that they are important in enzyme function. It has been demonstrated that the linker region of *C. japonicus* Xyn10A is essential for the activity of the enzyme against cellulose/hemicellulose complexes (Black *et al.*, 1996). The most likely explanation for the role of linkers is that they provide the catalytic module with the conformational freedom it requires to work efficiently while it is anchored to its substrate by the CBM (Black *et al.*, 1996).

CBM Family	Members	Representative	Protein size*	Ligand Specificity	Reference
1	170	*Trichoderma reesei* Cel7A	~ 40	Crystalline cellulose	Tomme *et al.*, 1995b
2a	137	*Cellulomonas fimi* Xyn10A	~ 100	Crystalline cellulose	McLean *et al.*, 2000; Simpson *et al.*, 2000
2b		*Cellulomonas fimi* Xyn11A		Xylan	Bolam *et al.*, 2001; Xie *et al.*, 2001a
3a	75	*Clostridium cellulovorans* CbpA	~ 150	Crystalline cellulose	Goldstein *et al.*, 1993
3b		*Clostridium thermocellum* Cel9I		Amorphous cellulose	Zverlov *et al.*, 2003
3c		*Clostridium thermocellum* Cel9N		Amorphous cellulose	Zverlov *et al.*, 2003
4a	19	*C. fimi* Cel9B	~ 150	Crystalline cellulose	Boraston *et al.*, 2002a
4b		*Rhodothermus marinus* Xyn10A		Xylan (preferentially)	Abou Hachem *et al.*, 2000; Simpson *et al.*, 2002; Abou Hachem *et al.*, 2002
5	134	*Erwinia chrysanthemi* Cel5	~ 60	Crystalline cellulose	Simpson and Barras, 1999
6	54	*C. thermocellum* Xyn11A	~ 120	Xylan (preferentially)	Czjzek *et al.*, 2001
8	2	*Dictyostelium discoideum* Cel9A	~ 150	Crystalline cellulose	Wang *et al.*, 2001
9	21	*Thermotoga maritima* Xyn10A	~ 170	Amorphous cellulose	Notenboom *et al.*, 2001
10	16	*Cellvibrio japonicus*** Xyn10A	~ 50	Crystalline cellulose	Ponyi *et al.*, 2000
11	4	*C. thermocellum* Cel26A	~ 180-200	Crystalline cellulose	Yague *et al.*, 1990
13	221	*Streptomyces lividans* Xyn10A	~ 150	Xylan	Boraston *et al.*, 2000b; Notenboom *et al.*, 2002; Schärpf *et al.*, 2002
15	2	*Cellvibrio japonicus*** Xyn10C	~ 110	Xylan	Szabó *et al.*, 2001
17	10	*Clostridium cellulovorans* Cel5A	~ 200	Cellulose	Boraston *et al.*, 2000a
22a	60	*Polyplastron multivesiculatum* Xyn10B	~ 160	Cellulose	Devillard *et al.*, 2003
22b		*C. thermocellum* Xyn10B		Xylan	Charnock *et al.*, 2000; Xie *et al.*, 2001b
23	1	*C. fimi* Man26A	~ 180	Mannan	Stoll *et al.*, 2000
24	6	*Penicillium purpurogenum* α-1,3-glucanase	~ 99	α-1,3-glucan	Fuglsang *et al.*, 2000
27	5	*Caldicellulosiruptor* (strain Rt8B.4) ManA		Mannan	Sunna *et al.*, 2001
28	10	*Bacillus* sp. Cel5A		Cellulose/ β-1,3-1,4-glucans	Boraston *et al.*, 2002b
29	1	*Piromyces equi* NCP1		Mannan/glucomannan	Freelove *et al.*, 2001; Charnock *et al.*, 2001
30	5	*C. thermocellum* CelJ		Cellulose	Arai *et al.*, 2003
31	3	*Alcaligenes* sp. 3XynAlc		β-1,3-xylan	Okazaki *et al.*, 2002
35	38	*C. japonicus* XynB *C. thermocellum* Man26A	~ 130	Xylan/mannan	Kellet *et al.*, 1990 Halstead *et al.*, 1999
36	5	*Paenibacillus polymyxa* XynD	~ 120-130	Xylan	Gosalbes *et al.*, 1991

Table 1.4: CBM families that bind cellulose and/or hemicellulose.

*Protein size denotes number of amino acids. ** denotes formerly *Pseudomonas cellulosa*. Based on data derived from CAZy (http://afmb.cnrs-mrs.fr/CAZY/.

1.4 Research objectives

1.4.1 Family 6 CBM

Recently, the X-ray crystal structure of a family 6 CBM from *Clostridium thermocellum* (*Ct*CBM6) was solved, enabling the rational design of site-directed mutagenesis (SDM) experiments to identify key residues involved in substrate binding. In conjunction with this novel structural data and previously reported results (Fernandes *et al.*, 1999), the overall aim of this study is to investigate the interaction between *Ct*CBM6 and its ligands. A site-directed mutagenesis approach was used to investigate the role of key amino acids in defining ligand specificity.

Specific objectives of this study are as follows:

1. To confirm the ligand specificity of *Ct*CBM6.
2. To investigate the affinity of *Ct*CBM6 for various soluble ligands and to determine the thermodynamic forces driving ligand binding.
3. To use SDM informed by the crystal structre of *Ct*CBM6 to dissect the functional significance of residues at the interacting surface.
4. To understand the structural basis for changes in affinity of selected mutants obtained in (3) and to identify the cleft used by *Ct*CBM6 in substrate binding.

1.4.2 Family 69 GH

It is apparent from a review of the literature regarding family 69 hyaluronidases that there remain lots of unanswered questions concerning their mode of action, the elucidation of the 3D structure and identification of key catalytic residues within the active site to further understand the ligand specificity and protein/ligand interactions of this family of enzymes. Since the complete genome sequence of *S. pyogenes* SF370 was published (Ferretti *et al.*, 2001), approximately 109 ORFs encoding proteins involved in carbohydrate metabolism could be identified, including enzymes involved in hyaluronan degradation. Hyaluronidases from *S. pyogenes* SF370 have been shown to be human virulence

factors (Hynes and Walton, 2000). Therefore, three putative hyaluronidases from *S. pyogenes* SF370 (ORFs HylP1, HylP2 and HylP3) and belonging to a novel family of GHs, for which no structural or mechanistic data is known, were cloned and biochemically characterised.

Specific aims of this study are as follows:

1. To clone, express and purify ORFs HylP1, HylP2 and HylP3 from *S. pyogenes* SF370, encoding putative hyalronidases.
2. To biochemically characterise HylP1, HylP2 and HylP3.
3. To identify the mode of action of HylP1, HylP2 and HylP3.
4. To crystallise and obtain X-ray diffraction data for HylP1, HylP2 and HylP3, which is pivotal to 3D structure determination.

2 Materials and Methods

2.1 Materials

2.1.1 Chemicals

The chemicals, enzymes and kits used throughout this study are listed in Appendix A. Unless otherwise stated, all solutions and buffers were made up with 18.2 MΩ H_2O purified using a Milli –Q Plus 185 water purification system, and stored at room temperature. For all molecular biology protocols, Sambrook and Russell (2001) was used as a general reference.

2.1.2 Reaction vessels and equipment

A list of reaction vessels and equipment used in this study can be found in Appendix D.

2.1.3 Media

Unless otherwise stated, all media were prepared using distilled water, and sterilised by autoclaving as detailed in section 2.2.1.1. Liquid media were stored at room temperature, and solid media at 4 °C. All adjustments to the pH of solutions are stated under the list of components, and were achieved using HCl or NaOH.

2.1.3.1 Liquid media

2.1.3.1.1 LB (Luria-Bertani) broth

Per L

Tryptone	10.0 g
Yeast extract	5.0 g
NaCl	10.0 g

pH 7.0

2.1.3.1.2 NZY Enrichment broth

Per L

NZ Amine (casein hydrolysate)	10.0 g
Yeast extract	5.0 g
NaCl	5.0 g
$MgSO_4.7H_2O$	2.0 g

pH 7.5

2.1.3.1.3 NZY^+ Enrichment broth

NZY^+ broth was used for the recovery of transformed chemically competent cells. The broth was made by the addition of 0.45 ml of supplement (stored at 4 °C) to 10 ml of NZY broth using sterile vessels and pipette tips.

NZY (supplement)

Per 45 ml

1 M $MgCl_2.6H_2O$	12.5 ml
1 M $MgSO_4.7H_2O$	12.5 ml
20 % (w/v) D-(+)-Glucose	20.0 ml

Filter sterilised.

2.1.3.1.4 Seleno-L-Methionine broth

Seleno-L-Methionine (Se-Met) broth was used for the production of Seleno-L-Methionine containing HylP1, HylP2 and HylP3 having been transformed into the methionine auxotroph *Escherichia coli* B834 (DE3).

All components were prepared fresh on day of use and filter sterilised, using Whatman filters 0.2 μm, except the $MgSO_4$ and M9 Salts, which were autoclaved prior to use (Section 2.2.1.1).

Per L

1 M $MgSO_4.7H_2O$	2.0 ml
20 X M9 Salts	100.0 ml
12.5 mg ml^{-1} $FeSO_4.7H_2O$	2.0 ml
40 % (w/v) D-(+)-Glucose	10.0 ml

4mg ml^{-1} Amino acid mix I	10.0 ml
4mg ml^{-1} Amino acid mix II	10.0 ml
1 mg ml^{-1} Vitamin mix	1.0 ml
10 mg ml^{-1} Seleno-L-methionine	4.0 ml

M9 Salts

Per L

NH_4Cl	20.0 g
KH_2PO_4	60.0 g
Na_2HPO_4	120.0 g

Amino Acid Mix I

Per 10 ml

Alanine	40.0 mg
Arginine	40.0 mg
Asparagine	40.0 mg
Aspartic acid	40.0 mg
Cysteine	40.0 mg
Glutamic acid	40.0 mg
Glutamine	40.0 mg
Glycine	40.0 mg
Histidine	40.0 mg
Isoleucine	40.0 mg
Leucine	40.0 mg
Lysine	40.0 mg
Proline	40.0 mg
Serine	40.0 mg
Threonine	40.0 mg
Valine	40.0 mg

Amino Acid Mix II

Per 10 ml

Phenylalanine	40.0 mg
Tryptophan	40.0 mg

Tyrosine	40.0 mg

Vitamin Mix

Per 1 ml	
Riboflavin	1.0 mg
Niacinamide	1.0 mg
Pyridoxine monohydrochloride	1.0 mg
Thiamine	1.0 mg

2.1.3.2 Solid media

2.1.3.2.1 LB agar

For the preparation of agar plates LB media was supplemented with the following.

Per 100 ml	
Agar (bacteriological agar N° 1)	2.0 g

Agar was then made soluble by autoclaving the media (Section 2.2.1.1), and poured into Petri dishes when the media had cooled to less than 55 °C.

2.1.3.3 Selective media

2.1.3.3.1 Antibiotics and other supplements

For antibiotic selection, appropriate volumes of stock solutions were added to liquid media that had been cooled to less then 55 °C. The composition of the stock solutions and final concentration used in the media are given in Table 2.1. Isopropylthio-β,D-galactoside (IPTG) was added to strains of *E. coli* carrying the *lacI*^{q} and *lacO* regions in the plasmid only, or in the plasmid and the bacterial genome to induce transcription of the recombinant gene. IPTG (100 mM; 24 mg ml^{-1}) stock dissolved in 18.2 MΩ H_2O) was added to liquid media to a final concentration of 1.0 mM (240 μg ml^{-1}). 5-bromo-4-chloro-3-indoyl-β-D-galactoside (X-GAL; 40 mg ml^{-1} stock dissolved in dimethyl formamide) was added to liquid media, to detect the insertional inactivation of the β-galactosidase gene, to a final concentration of 80 μg ml^{-1}.

Antibiotic	Stock concentration	Working concentration	Solvent	Storage
Ampicillin	10 mg ml^{-1}	100 µg ml^{-1}	18.2 MΩ H_2O	4 °C for less than 7 days or – 20 °C
Kanamycin	10 mg ml^{-1}	50 µg ml^{-1}	18.2 MΩ H_2O	– 20 °C
Tetracycline	12.5 mg ml^{-1}	12.5 µg ml^{-1}	50 % Ethanol	– 20 °C

Table 2.1: Stock concentrations and final concentrations of antibiotics in media.

2.1.3.3.2 Blue/white selection

For the detection of recombinant plasmids in pGEM®-T Easy through insertional inactivation of the β-galactosidase gene, LB agar previously supplemented with ampicillin (Section 2.1.3.3.1) was also supplemented with IPTG and X-GAL. The stocks were then diluted to the appropriate concentration in the media (Section 2.1.3.3.1).

2.1.4 Cryogenic storage of bacterial stocks

Bacterial stocks were created by mixing 0.5 ml of the culture broth with 0.5 ml of 50 % (v/v) glycerol. These stocks were then stored at –80 °C.

Per 100 ml

100 % (v/v) Glycerol 50.0 ml

2.1.5 Bacterial strains and plasmids

All bacterial strains were maintained at –80 °C in 25 % (v/v) glycerol, and were used as detailed in Table 2.2. Furthermore, every month, bacterial stocks of *E. coli* XL1-Blue, XL10-Gold and BL21 (DE3) were streaked (using a sterile wire loop) onto an agar plate supplemented with the appropriate antibiotic (if any). After growth at 37 °C overnight, the plate was stored at 4 °C for 1 month. This precaution was taken to ensure bacterial stocks were not contaminated, and provide a re-inoculum in the event of a stock becoming contaminated. The restriction maps for the vectors used in this study are given in Appendix B.

E. coli strain/Vector	Characteristics	Application	Reference/ Supplier
XL1-Blue	*recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac* [F' *proAB lacI^{q}ZΔM15* Tn*10* (Tet^r)]	Cloning host	Bullock *et al.*, 1987 /Stratagene
XL10-Gold	Tet^R Δ *(mcrA) 183* Δ *(mcrCB-hsd SMR-mrr) 173 endA1 supE44 thi-1 recA1 gyrA96 relA1 lac Hte* [F' *proAB lacI^{q}ZΔM15* Tn*10* (Tet^r) *Amy* Cam^R]	Cloning host	Bullock *et al.*, 1987 /Stratagene
One Shot™ TOP10	F$^-$ *mcrA* Δ (*mrr-hsd*RMS-*mcr*BC) ϕ80*lacZ*ΔM15 Δ*lacX74 deoR recA1 araD139* Δ(*ara-leu*)7697 *galU galK rspL endA1 nupG*	Cloning host	Invitrogen Corp.
BL21 (DE3)	F$^-$*ompT hsdS$_B$(r$_B^-$m$_B^-$) gal dcm* (DE3)	Expression host	Studier *et al.*, 1986 /Novagen
B834 (DE3)	F$^-$*ompT hsdS$_B$(r$_B^-$m$_B^-$) gal dcm met* (DE3)	Expression host (recombinant protein lacking methionine)	Novagen
pGEM®-T Easy	Amp^r, T7 *lacZ*	Cloning vector	Promega
pCR®-Blunt	Kan^r, Zn^r, T7 *lacZα-ccdB*	Cloning vector	Invitrogen Corp.
pET-21a	Amp^r, T7 *lac*, *lacI^q*	Expression vector	Novagen
pET-22b	Amp^r, T7 *lac*, *lacI^q*	Expression vector	Novagen
pET-28a	Kan^r, T7 *lac*, *lacI^q*	Expression vector	Novagen

Table 2.2: Bacterial strains and plasmids used in this study.

2.1.6 Molecular biology chemicals, enzymes and kits

2.1.6.1 Agarose gel electrophoresis

2.1.6.1.1 TAE Running buffer (50 x stock)

Per L	
Tris-HCl	242.0 g
17.51 M Glacial acetic acid	57.1 ml
0.5 M EDTA pH 8.0	100.0 ml

This buffer was diluted to 1 x by a 1:50 dilution prior to use in the preparation of agarose gels and use as reservoir buffer in gel tanks.

2.1.6.1.2 Bromophenol blue (6 x) sample loading buffer

Per 10 ml	
Bromophenol blue	0.025 g
Sucrose	4.000 g

This buffer was diluted to 1 x with the sample prior to loading on an agarose gel.

2.1.6.1.3 Size Standards

Agarose gel size standards were diluted to a concentration of 1 µg/12 µl prior to loading on a gel. Three standard mixes were used, a 1 kb DNA ladder (10, 8, 6, 5, 4, 3, 2, 1.5, 1, 0.5 kb), a 100 bp DNA ladder (1500, 1000, 900, 800, 700, 600, 500, 400, 300, 200, 100 bp) and lambda DNA / *Hin*d III digest (23130, 9416, 6557, 4361, 2322, 2027, 564, 125 bp).

2.1.6.1.4 Ethidium bromide

This chemical was used diluted 1:1000 prior to use from an original concentration of 10 mg ml^{-1} to a working concentration of 10 µg ml^{-1}.

2.1.6.2 General use chemicals

2.1.6.2.1 TE buffer

Per L	
0.5 M Tris-HCl pH 7.5	20.0 ml
0.5 M EDTA pH 8.0	2.0 ml

pH 7.5

2.1.6.2.2 Crude plasmid preparation buffer (STETS buffer)

Per L	
Sucrose	80.0 g
Triton X-100	50.0 ml
0.5 M EDTA pH 8.0	100.0 ml
Tris-HCl	6.06 g

pH 8.0

2.1.6.2.3 FSB (Transformation solution)

Per 100 ml	
1 M Potassium acetate pH 7.5 (pH adjusted with acetic acid)	1.0 ml
$MnCl_2.4H_2O$	890.0 mg
$CaCl_2.2H_2O$	150.0 mg
KCl	750.0 mg
Hexaamminecobalt chloride	80.0 mg
100 % (v/v) Glycerol	10.0 ml

The solution was then adjusted to pH 6.4 using HCl only (no readjustment with base), and topped up to 100 ml using 18.2 MΩ H_2O. Finally, the solution was filter sterilised and stored at 4 °C.

2.1.6.3 Kits

The molecular biology kits employed in this study are detailed in Table 2.3. All kits and components were stored at room temperature, except the cell re-suspension buffer of plasmid purification kits, which was maintained at 4 °C.

Kit	Application
NucleoSpin® Plasmid	Purification of recombinant plasmids
NucleoSpin® Extract 2 in 1	Purification of PCR products and some restriction enzyme digests. Agarose gel purification of DNA bands.

Table 2.3: Molecular biology kits and applications.

For a list of components of each kit, refer to Appendix A5.

2.1.6.4 Enzymes

All enzyme buffers and reaction co-constituents were maintained at – 20 °C, in 50 % (v/v) glycerol. A list of general enzymes used in this study is shown in Table 2.4. However, a full list of all enzymes and constituents of reaction buffers is in Appendix A3.

Enzyme (stock concentrations)	Co-constituents
Pfu DNA polymerase (3 U μl^{-1})	Reaction buffer (1 x): 10 mM KCl, 10 mM $(NH_4)_2SO_4$, 20 mM Tris-HCl pH 8.8, 2 mM $MgSO_4$, 0.1 % Triton® X-100, 0.1 mg ml^{-1} BSA
Platinum® *Pfx* DNA polymerase (2.5 U μl^{-1})	Amplification buffer (1 x): Patented
Taq DNA polymerase (5 U μl^{-1})	Reaction buffer (1 x): 10 mM Tris-HCl pH 9.0, 1.5 mM $MgCl_2$, 50mM KCl, 0.1 % Triton® X-100
Restriction enzymes: *Bam*H I, *Eco*R I, *Eco*R V, *Hin*d III, *Nde* I, *Sal* I, *Xho* I	A list of buffer constituents used for each restriction enzyme is given in Appendix A3
RNase (10 mg ml^{-1})	Added to Elution Buffer: 10 mM Tris-HCl pH 7.5, 1 mM EDTA, 100 μg ml^{-1} RNase
Alkaline phosphatase (Calf intestinal)	Reaction buffer (1 x): 50 mM Tris-HCl pH 9.3, 1 mM $MgCl_2$, 0.1 mM $ZnCl_2$, 1 mM spermidine
T_4 DNA ligase (133 U μl^{-1})	Ligation buffer (1 x): 50 mM Tris-HCl pH 7.5, 10 mM $MgCl_2$, 10 mM DTT, 1 mM ATP, 25 μg ml^{-1} BSA
Lysozyme (Hen egg white; 10 mg ml^{-1})	Used unbuffered
Dpn I Restriction enzyme (10 U μl^{-1})	Used unbuffered

Table 2.4: Enzymes and co-constituents.

2.1.7 Primer Sequences

2.1.7.1 Creation of *Ct*CBM6 mutants

The primer sequences given in Table 2.5 were used for the construction of site-directed mutants. Mutagenic primers were designed such that the sequence of bases either side of the mutated base was fully complementary to the sense and antisense strands of wild type *Ct*CBM6.

2.1.7.2 For amplification of HylP1, HylP2 and HylP3

The primer sequences given in Table 2.6 were used for the amplification of HylP1, HylP2 and HylP3.

2.1.8 Vectors

For use in the cloning of PCR products, pGEM®-T Easy and pCR®-Blunt were used, and for the subsequent expression of cloned open reading frames pET-21a, pET-22b, and pET-28a (Table 2.2) were used. Restriction maps of the multiple cloning regions of the stated vectors are provided in Appendix B.

2.1.9 Solutions, resins and buffers for protein analysis and purification

2.1.9.1 SDS – PAGE chemicals

2.1.9.1.1 12 % Acrylamide resolving gel components

40 % (v/v) solution (37.5:1 acrylamide: bisacrylamide)	3.0 ml
Solution B*	2.5 ml
18.2 MΩ H_2O	4.5 ml
10 % (w/v) Ammonium persulphate	50 µl
TEMED	10 µl

*Solution B

Per 100 ml

2 M Tris-HCl pH 8.8	75 ml
10 % (w/v) SDS	4 ml

4% Acrylamide stacking gel components

40 % (v/v) solution (37.5:1 acrylamide: bisacrylamide)	0.5 ml
Solution C*	1.0 ml
18.2 MΩ H_2O	2.5 ml
10 % (w/v) Ammonium persulphate	30 µl
TEMED	10 µl

Primer name	Primer sequence	Primer length (mer)	T_m (°C)	%GC content	% Mismatch
N120Af	5'ggTATTCTCAggTCCTgTT**gCg**ATTgACTACTTCATATTCgACTC3'	45	77.1	42.2	6.7
N120Ar	3'gAgTCgAATATgAAgTAgTCAAT**CgC**AACAggACCTgAgAATACC5'	45	77.1	42.2	6.7
Y34Af	5'CggAggAAgCggTATAggT**gCg**ATTgAAAgCggTgACTATC3'	41	79.8	53.7	7.3
Y34Ar	3'gATAgTCACCgCTTTCAAT**CgC**ACCTATACCgCTTCCTCCG5'	41	79.8	53.7	7.3
R72Mf	5'CACCAATATCCAgTTAA**Tg**CTCggAAgCCCgACCg3'	35	78.8	54.3	5.7
R72Mr	3'gTggTTATAggTCAAT**TAC**gAgCCTTCgggCTggC5'	35	78.8	54.3	5.7
Y40Mf	5'gCggTATAggTTATATTgAAAgCggTgAC**ATg**CTggTATTTAAC3'	43	76.0	41.9	7.0
Y40Mr	3'CgCCATATCCAATATAACTTTCgCCACTg**TAC**gACCATAAATTg5'	43	76.0	41.9	7.0
Y112Mf	5'CAggACAgCACgACTTAA**Tg**CTggTATTCTCAggTCCTg3'	39	77.5	51.3	7.7
Y112Mr	3'gTCCTgTCgTgCTgAAT**TAC**gACCATAAgAgTCCAggAC5'	39	77.5	51.3	7.7
W92Mf	5'CggTggCTTCCACAggTggT**ATg**AACAATTACgAgg3'	36	78.8	52.8	5.6
W92Mr	3'gCCACCgAAggTgTCCACCA**TAC**TTgTTAATgCTCC5'	36	78.8	52.8	5.6

Table 2.5: Mutant primer sequences for *Ct*CBM6.

Letters in embolden denote the positions in the gene sequence that were mutated. The template DNA was p*Ct*Xyn11A. Primer length is the number of bases in the primer. T_m (melting temperature) is specific for each primer and was calculated to be ≥ 76 °C from the equation: $T_m = 81.5 + 0.41$ (% GC) – 675 / N - % mismatch (where % GC is the percentage of guanine and cytosine residues in the primer; N is the primer length in bases, and % mismatch is the percentage of the primer non complementary at the point of mutation).

Primer name	Primer sequence
SF370HylP1f	5'CATATgAgTgAAAATATACCgCTg3'
SF370HylP1/3r	5'CTCgAgTTTTTTTAgTATgAgTTTTTTTAAC3'
SF370HylP1/3rstop	5'ggATCCCTATTTTTTTAgTATgAg3'
SF370HylP15'	5'ATggCAATCAATgggTCAAg3'
SF370HylP35'	5'gAAgCTAATgCAgAggTCAAAg3'
SF370HylP2/3f	5'CATATggCTgAAAATATACCgC3'
SF370HylP2r	5'CTCgAgCATTTgCTTAgCCgCTAAg3'
SF370HylP2rstop	5'CTCgAgTTACATTTgCTTAgCCgC3'

Table 2.6: Primer sequences used for the amplification of HylP1, HylP2 & HylP3 from glycoside hydrolase family 69.
The genomic DNA used as the template was *Streptococcus pyogenes* SF370 (ATCC 700294).

Primer pair	PCR product amplified	Primary (1°) annealing temperature (°C)	Secondary (2°) annealing temperature (°C)
HylP15' + HylP1/3rstop	HylP1 upstream of gene (Stop)	37.3	50.9
HylP15' + HylP1/3r	HylP1 upstream of gene (No stop)	44.9	53.9
HylP2/3f + HylP2rstop	HylP2 gene (Stop)	46.4	51.5
HylP2/3f + HylP2r	HylP2 gene (No stop)	47.4	51.5
HylP35' + HylP1/3rstop	HylP3 upstream of gene (Stop)	37.3	50.9
HylP35' + HylP1/3r	HylP3 upstream of gene (No stop)	44.9	53.4
HylP1f + HylP1/3rstop	HylP1 gene (Stop)	37.3	50.9
HylP1f + HylP1/3r	HylP1 gene (No stop)	44.9	52.6
HylP2/3f + HylP1/3rstop	HylP3 gene (Stop)	37.3	50.9
HylP2/3f + HylP1/3r	HylP3 gene (No stop)	44.9	51.5

Table 2.7: Primary and secondary annealing temperatures used for PCR amplification of HylP1, HylP2, and HylP3 genes from glycoside hydrolase family 69.
The template DNA used to PCR HylP1 upstream, HylP2, and HylP3 upstream was *S. pyogenes* SF370 gDNA (ATCC 700294). For the specific amplification of genes HylP1 and HylP3, pGEM®-T cloned DNA from PCR products of HylP1 upstream and HylP3 upstream were used as template. The T_m for each primer pair was calculated from the equation: $T_m = 69.3 + (0.41 \times \% GC) - (650 / \text{primer length})$.

*Solution C

Per 100 ml

1 M Tris-HCl pH 6.8	50 ml
10 % (w/v) SDS	4 ml

2.1.9.1.2 15 % Acrylamide resolving gel components

40 % (v/v) solution (37.5:1 acrylamide: bisacrylamide)	3.75 ml
Solution B*	2.50 ml
18.2 MΩ H_2O	3.75 ml
10 % (w/v) Ammonium persulphate	50 µl
TEMED	10 µl

*Solution B

Per 100 ml

2 M Tris-HCl pH 8.8	75 ml
10 % (w/v) SDS	4 ml

4% Acrylamide stacking gel components

40 % (v/v) solution (37.5:1 acrylamide: bisacrylamide)	0.5 ml
Solution C*	1.0 ml
18.2 MΩ H_2O	2.5 ml
10 % (w/v) Ammonium persulphate	30 µl
TEMED	10 µl

*Solution C

Per 100 ml

1 M Tris-HCl pH 6.8	50 ml
10 % (w/v) SDS	4 ml

2.1.9.1.3 SDS - PAGE running buffer 10 x (Stock)

Running buffer was made up at 10 x stock concentration and diluted 1:10 before use.

Per L

Tris-HCl	30.3 g
Glycine	144.0 g
SDS	10.0 g

The pH of the Tris-HCl and glycine was adjusted to 8.8 in a volume of ~900 ml prior to the addition of SDS.

2.1.9.1.4 SDS - PAGE sample buffer

Per 10 ml

60 mM Tris-HCl pH 6.8	0.6 ml
50 % (v/v) Glycerol	5.0 ml
10 % (w/v) SDS	2.0 ml
14.4 mM β-Mercaptoethanol	0.5 ml
1 % (w/v) Bromophenol blue	1.0 ml

Stored at –20 °C in 0.5 ml aliquots.

2.1.9.1.5 Solublising SDS - PAGE sample buffer

Per 10 ml

SDS - PAGE sample buffer	7.6 ml
Urea	2.4 g

Stored at 4 °C.

2.1.9.1.6 Protein size standards

High molecular weight range (M.W. 36, 45, 55, 66, 84, 97, 116 and 205 kDa)
Low molecular weight range (M.W. 20, 24, 29, 36, 45 and 66 kDa)

Lyophilised standards were reconstituted with 100 µl of 18.2 MΩ H_2O to give a final concentration of ~2.0-3.5 mg ml^{-1}, which was aliquoted into 4 µl amounts and stored at –20 °C. Refer to Appendix A4 for a list of proteins used to produce the appropriate sized bands.

2.1.9.1.7 Coomassie blue gel stain solution

Per L

Coomassie Blue R-250	1.0 g
Acetic acid (Glacial)	100.0 ml
Methanol	450.0 ml

2.1.9.1.8 Coomassie gel destain solution

Per L

Acetic acid (Glacial)	100.0 ml
Methanol	100.0 ml

2.1.9.2 Affinity Gel Electrophoresis (AGE) chemicals

2.1.9.2.1 7.5 % Continuous native PAGE components

40 % (v/v) solution (37.5:1 acrylamide: bisacrylamide)	1.88 ml
Non-denaturing gel buffer (10 x)	1.00 ml
18.2 MΩ H_2O	6.06 ml
10 % (w/v) Ammonium persulphate	50.0 µl
TEMED	10.0 µl

The above solution was then gently mixed and divided into two, where 1 ml of a 1 % (w/v) ligand solution was added, or 1 ml 18.2 MΩ H_2O was added as a control.

2.1.9.2.2 Protein standard

Bovine serum albumin (BSA) was run on affinity gels as a protein control standard. BSA was made up at a stock concentration of 2.5 mg ml^{-1} in sterile 18.2 MΩ H_2O and was then diluted 1:1 with loading buffer to give a final concentration of 1.25 mg ml^{-1}.

2.1.9.2.3 Non-denaturing running buffer 10 x (Stock)

Running buffer was made up at a 10 x stock concentration and diluted 1:10 before use.

Per L

Tris-HCl	30.28 g
Glycine	187.75 g

pH 8.3

2.1.9.2.4 Non-denaturing loading buffer

Per 10 ml

10 % (v/v) Glycerol	9.0 ml
10 x Non-denaturing gel buffer	1.0 ml
Bromophenol Blue	0.0005 g

2.1.9.3 Polyacrylamide gel electrophoresis (PAGE) analysis chemicals

2.1.9.3.1 15.5 % Continuous native PAGE components

40 % (v/v) solution (37.5:1 acrylamide: bisacrylamide)	9.69 ml
10 x TBE buffer	2.50 ml
18.2 MΩ H_2O	12.66 ml
10 % (w/v) Ammonium persulphate	125.0 µl
TEMED	25.0 µl

2.1.9.3.2 Alcian Blue stain

Per L

Alcian Blue GX8	0.5 g

2.1.9.3.3 Silver stain reagents

Oxidation step

Per L

Potassium dichromate	1.0 g
0.0032 N Nitric Acid	0.2 ml

Silver stain

Per L

Silver nitrate	2.04 g

Image developer solution

Per L

Sodium Carbonate	29.68 g
100 % (v/v) Formaldehyde	0.5 ml

Termination step

Per L

100 % (v/v) Glacial acetic acid	10 ml

2.1.9.3.4 Tris-Borate-EDTA (TBE) buffer 10 x (stock)

Per L

Tris-HCl	121.1 g
Boric Acid	61.83 g
EDTA	3.72 g

pH 8.3

2.1.9.3.5 PAGE sample buffer

Per 10 ml

Sucrose	6.85 g
10 x TBE buffer	1.0 ml

2.1.9.3.6 PAGE tracking dye

Per 10 ml

1 % (w/v) Bromophenol blue	50 µl
Sucrose	1.03 g
10 x TBE buffer	1.0 ml

2.1.9.4 Protein purification

2.1.9.4.1 Periplasmic cell resuspension buffer (post-induction)

Per L

Sucrose	200.0 g
Tris-HCl	6.06 g

pH 8.0

2.1.9.4.2 Cell resuspension buffer (post-induction)

Per L

HEPES	4.77 g
NaCl	29.22 g
Imidazole	0.68 g

pH 7.4

If the recombinant protein had been expressed in *E. coli* B834 (DE3), 1 mM β-mercaptoethanol (final concentration) was added to the cell resuspension buffer.

2.1.9.4.3 Purification resins

Affinity purification resin:

Talon™ Metal Affinity resin

Chelating Sepharose™ Fast Flow resin

Gel filtration resin:

Superdex 200 prep grade

Buffer exchange resin:

Sephadex™ G-25 medium

All resins were stored in 20 % (v/v) ethanol.

2.1.9.4.4 Talon™ affinity purification buffers (Gravity flow)

Prior to loading the periplasmic fraction, *Ct*CBM6 wild type was conditioned by the addition of the following components (the result was equivalent concentrations in the load material and wash buffer):

Per 10 ml of periplasmic fraction:

Tris-HCl	0.024 g
NaCl	0.058 g

Wash buffer

Per L

Tris-HCl	2.42 g

NaCl	5.84 g

pH 8.0

Elution buffer

Per L	
Tris-HCl	2.42 g
NaCl	5.84 g
Imidazole	6.80 g

pH 8.0

2.1.9.4.5 Ni^{2+} affinity purification buffers (Gravity flow)

Start buffer

Per L	
Na_2HPO_4	2.84 g
NaCl	29.2 g
Imidazole	0.68 g

pH 7.4

Wash buffer

Per L	
Na_2HPO_4	2.84 g
NaCl	29.2 g
Imidazole	6.80 g

pH 7.4

Elution buffer

Per L	
Na_2HPO_4	2.84 g
NaCl	29.2 g
Imidazole	34.0 g

pH 7.4

2.1.9.4.6 Ni^{2+} affinity purification buffers (Automated purification, gradient flow)

Solution A:

Per L	
HEPES	4.77 g
NaCl	29.22 g
Imidazole	0.68 g

pH 7.4

Solution B:

Per L	
Imidazole	68.0 g

pH 7.4

2.1.9.4.7 Gel filtration buffers

Per L	
HEPES	4.77 g
$CaCl_2.2H_2O$	29.4 g

pH 7.4

If the recombinant protein had been expressed in *E. coli* B834 (DE3), 10 mM dithiothreitol (DTT; final concentration) was added to the gel filtration buffer.

2.1.9.4.8 Buffer exchange buffers (Post-gel filtration)

This was achieved either using 10 kDa concentrators or PD-10 columns.

For gel filtration:

Per L	
HEPES	4.77 g

$CaCl_2.2H_2O$ 29.4 g

pH 7.4

For crystallisation:

Per L

HEPES 2.38 g

$CaCl_2.2H_2O$ 7.35 g

pH 7.4

If the recombinant protein had been expressed in *E. coli* B834 (DE3), 10 mM or 1 mM DTT (final concentration) was added to the buffers, respectively.

2.1.9.5 Dialysis

Dialysis of proteins or substrates was achieved in cellulose membrane tubing (size 25 mm x 16 mm; retains > 90 % Cytochrome C (M.W. 12 400 Da); Sigma) and secured with dialysis tubing closures (gripping length 50 mm; Sigma).

Purified *Ct*CBM6 wild type and mutant proteins were dialysed against 3 x 8 L:

Dialysis buffers

Per L

NaH_2PO_4 6.9 g

pH 7.0

Sodium hyaluronate, soluble oat spelt and birchwood xylans were dialysed against 18.2 MΩ H_2O (3 x 4 L).

2.1.9.5.1 Preparation of dialysis tubing

The dialysis tubing was cut into pieces of convenient length and boiled for 10 min in a large volume of 18.2 MΩ H_2O. From this point onward, the tubing was always handled with gloves. The tubing was then rinsed thoroughly inside and out with 18.2 MΩ H_2O and once cool, was then stored at 4 °C in 18.2 MΩ H_2O.

2.1.9.6 DNSA reagents

Per L	
3,5-Dinitrosalicylic Acid	10.0 g
Phenol	2.0 g
NaOH	10.0 g

Per 10 ml	
Sodium sulphite	0.5 g

Per 10 ml	
Glucose	2.0 g

2.1.9.7 Protein crystallisation screens

The solutions used to screen for crystal development are given in Appendix C.

2.1.9.8 Reagents for Bradfords assay

BSA was made at a concentration of 10 mg ml^{-1} and stored at –20 °C, for subsequent dilution for assays. Bradford Reagent (Brilliant blue G in phosphoric acid and methanol) was stored at 4 °C.

2.1.10 Substrates

Soluble plant structural polysaccharides were used in qualitative AGE (Table 2.8) and made up at a stock concentration of 1 % (w/v) in sterile 18.2 MΩ H_2O and used at a final concentration of 0.1 % (w/v). Quantitative AGE was also carried out in the presence of a range of oat spelt xylan concentrations from 0.005 – 1.0 mg ml^{-1}.

Soluble plant structural oligosaccharides and polysaccharides were used in isothermal titration calorimetry (ITC; Table 2.8) and made up at stock concentrations of 5-30 mM and 2-4 mg ml^{-1}, respectively, in 50 mM sodium phosphate buffer, pH 7.0.

Polysaccharide/oligosacchride	AGE	ITC
Apple pectin	X	
Barley β-glucan	X	
Birchwood xylan	X	X
Carboxymethylcellulose	X	
Carobgalactomannan	X	
CM-Pachyman	X	
Hydroxyethly cellulose	X	
Lime pectin	X	
Locust bean galactomannan	X	
Methylcellulose	X	
Oat spelt xylan	X	X
Pectin galactan	X	
Potato galactan	X	
Rhamnogalacturonan	X	
Rye arabinoxylan	X	X
Sugar beet arabinan	X	
Wheat arabinoxylan	X	X
Xylose (X_1)		X
Xylobiose (X_2)		X
Xylotriose (X_3)		X
Xylotetraose (X_4)		X
Xylopentaose (X_5)		X
Xylohexaose (X_6)		X
Cellopentaose (G_5)		X
Cellohexaose (G_6)		X

Table 2.8: Plant structural polysaccharides and oligosaccharides used in AGE and ITC.

Soluble sodium hyaluronate was used in DNSA assays and made up at stock concentrations of 4.33, 3.0 and 2.66 mg ml^{-1} in sterile 18.2 MΩ H_2O and used at a final concentration of 4.0, 2.66 and 2.0 – 0.5 mg ml^{-1}, respectively.

2.1.10.1 Preparation of xylan fractions

Soluble and insoluble fractions of oat spelt xylan and birchwood xylan were prepared as described by Ghangas *et al.* (1989). Oat spelt xylan or birchwood xylan (20 g) was resuspended in 200 ml 18.2 MΩ H_2O, the pH adjusted to 10.0 with 1 M NaOH, and stirred for 1 h at room temperature. To neutralise the solution, 1 M acetic acid was added until the pH was 7.0 and the soluble and insoluble fractions were separated by centrifugation at 10 000 x g, at room temperature for 10 min. The supernatant was decanted and dialysed three times in 4 L 18.2 MΩ H_2O (Section 2.1.9.5) overnight, frozen at – 80 °C, lyophilised (Section 2.1.2) and stored at room temperature for future use. The insoluble pellet

was resuspended in 200 ml 18.2 MΩ H_2O and centrifuged at 10 000 x g for 10 min, before being washed several times again with 18.2 MΩ H_2O. The washed insoluble xylan was then frozen at – 80 °C, lyophilised (Section 2.1.2) and stored at room temperature for future use.

2.1.10.2 Preparation of dialysed sodium hyaluronate

Sodium hyaluronate was prepared in the presence of the metal ion chelator EDTA. Sodium hyaluronate (1 g) was added to 100 ml 18.2 MΩ H_2O and autoclaved (Section 2.2.1.1). This solution was then stirred for 5 h at room temperature with 50 mM EDTA, pH 8.0. Precipitates were then removed by centrifugation for 15 min at 14 000 x g and supernatants were subsequently dialysed three times in 4 L 18.2 MΩ H_2O (Section 2.1.9.5) overnight. The dialysed substrate was then aliquotted in 20 ml volumes into Petri dishes, frozen at –20 °C, lyophilised (Section 2.1.2) overnight and stored at room temperature for future use. Prior to kinetic analysis, the substrate was resuspended in 18.2 MΩ H_2O.

2.2 Methods

2.2.1 Microbiology methods

2.2.1.1 Sterilisation

Unless stated otherwise, all solutions and apparatus were sterilised by autoclaving (Appendix D1) at 121 °C, with a pressure of 1.05 bar for 20 min. Solutions that could not be autoclaved were filter sterilised through a 0.2 µm Ministart® filter unit (Sartorius), attached to a suitable sterile syringe (Plastipak®, Becton Dickinson).

2.2.1.2 Growth of bacteria harbouring plasmids

2.2.1.2.1 Growth of bacteria for standard/crude plasmid purification

E. coli strain XL1-Blue was grown in 28 ml universals containing 3 ml of LB media (Section 2.1.3.1.1) or *E. coli* strain XL10-Gold was grown in 250 ml conical flasks containing 50 ml of LB media, supplemented with the appropriate antibiotic, the resistance gene for which was encoded on the plasmid to be

purified. The cultures were inoculated from an agar plate, glycerol stock or liquid culture using a sterile wire loop, and grown overnight at 37 °C with orbital shaking at 200 rpm.

2.2.1.2.2 Growth of bacteria for starter cultures

E. coli strains XL1-Blue, XL10-Gold, BL21 (DE3), or B834 (DE3) were grown in 28 ml sterile glass universals containing 5 ml of LB media (Section 2.1.3.1.1) supplemented with the appropriate antibiotic, the resistance gene for which is encoded on the plasmid carried by the cells. The cultures were inoculated from an agar plate, glycerol stock or liquid culture using a sterile wire loop, and grown overnight at 37 °C with orbital shaking at 150 rpm.

2.2.1.2.3 Growth of bacteria for the expression and subsequent purification of protein

A starter culture of *E. coli* BL21 (DE3) was grown as described in section 2.2.1.2.2, and was added at a 1 % (v/v) ratio to a conical flask containing LB media (Section 2.1.3.1.1) supplemented with the same antibiotic used in the starter culture. The volume of LB media used depended on the amount of protein required to perform appropriate biochemical or structural techniques. Table 2.9 shows the three scales of expression used.

Scale of expression	Volume of LB media	Size of conical flask	Volume of starter culture added
Small	50 ml	250 ml	0.5 ml
Medium	100 ml	500 ml	1.0 ml
Large	1000 ml	2500 ml (Baffled)	10.0 ml

Table 2.9: Volumes and vessels used in the growth of expression strains.

After addition of the starter, the culture was grown at 37 °C with orbital shaking at 200 rpm until the culture reached an OD_{600} of ~0.6. The OD of the culture was checked by transferring 1 ml of culture to a cuvette and reading the OD in a spectrophotometer previously zeroed with LB media and set to 600 nm. Once at 0.6, the cells were induced by the addition of 1 % (v/v) 100 mM IPTG (Section 2.1.4.3.1). Post induction, the culture was incubated at 30 °C overnight with orbital shaking at 100 rpm.

2.2.1.2.4 Growth of bacteria for the expression and subsequent purification of Se-Met proteins

A starter culture of *E. coli* B834 (DE3) was grown as described in section 2.2.1.2.2, and was added at a 1 % (v/v) ratio to a conical flask containing 50 ml LB media (Section 2.1.3.1.1) in a 500 ml conical flask supplemented with the same antibiotic used in the starter culture. After addition of the starter, the culture was grown at 37 °C with orbital shaking at 200 rpm until the culture reached an OD_{550} of ~0.2. The OD of the culture was checked by transferring 1 ml of culture to a cuvette and reading the OD in a spectrophotometer previously zeroed with LB media and set to 550 nm. Once at 0.2, the cells were harvested by centrifugation at 4 °C for 10 min at 4000 x g and the supernatant discarded. The pellet was kept on ice while being resuspended with 10 ml Se-Met media (Section 2.1.3.1.4). The resuspended cells were then centrifuged at 4 °C for 10 min at 4000 x g and the process repeated twice more. Once the cells (10 ml) had been washed three times, this was all used to inoculate the remaining Se-Met media (~ 970 ml) in a 2.5 L baffled flask. This culture was then grown at 37 °C with orbital shaking at 200 rpm until the culture reached an OD_{550} of ~0.6. The OD of the culture was checked by transferring 1 ml of culture to a cuvette and reading the OD in a spectrophotometer previously zeroed with Se-Met media (that had been removed before inoculation) and set to 550 nm. Once at 0.6, the cells were induced by the addition of 1 % (v/v) 100 mM IPTG (Section 2.1.4.3.1). Post induction the culture was incubated at 30 °C overnight with orbital shaking at 100 rpm.

2.2.1.3 Plating bacteria

Prior to plating, LB agar plates were surface dried by placing open face down at 65 °C for 10 min. A glass spreader was sterilised by immersion in 70 % (v/v) ethanol, followed by removal of excess ethanol by passing quickly through a blue Bunsen flame, and used to spread 50-200 µl of bacterial suspension evenly over the surface of the agar. When all the liquid had been absorbed into the agar, the plate was incubated inverted at 37 °C for ~ 16 h.

2.2.1.4 Preparation of chemically competent cells

For all strains of *E. coli*, chemically competent cells were freshly prepared using the procedure detailed here and was essentially by the method of Hanahan (1983).

Firstly, a starter culture was inoculated from an agar plate containing the required strain of *E. coli*, and in the case of XL1-Blue and XL10-Gold cells, the starter was supplemented with tetracycline (Section 2.1.4.3.1). In the case of BL21 (DE3) and B834 (DE3), no antibiotic was required. After overnight growth, the starter culture was used to inoculate 50 ml of LB media in a 250 ml conical flask by the addition of 1 % (v/v) starter. The culture was then allowed to grow at 37 °C with shaking at 200 rpm up to an OD_{550} of 0.55 (approximately 2 hours); at this point the culture was placed on ice for 30 min. After incubation on ice, the culture was pelleted by centrifugation at 2000 x g for 10 min at 4 °C and the supernatant discarded. The cell pellet was then re-suspended (by gentle vortex) in 4 ml of ice-cold FSB solution (Section 2.1.6.2.3), and the cells incubated on ice for a further 15 min, prior to re-centrifugation at 2000 x g for 10 min at 4 °C and rejection of the supernatant. The cell pellet was re-suspended (by gentle vortex) in 720 µl of ice-cold FSB solution and 26 µl of DMSO was then added in a drop-by-drop fashion. The cells were then incubated on ice for a further 15 min prior to the addition of another 26 µl of DMSO in a drop-by-drop fashion. Finally, the cells were placed on ice, and at this point XL1-Blue and XL10-Gold cells were ready to use. However, owing to the fact that transformations into BL21 (DE3) and B834 (DE3) do not require highest efficiencies, these cells were able to be aliquoted into 50 µl volumes and snap frozen by submersion in liquid nitrogen for storage at –80 °C.

2.2.2 DNA Methods

2.2.2.1 General DNA methods

2.2.2.1.1 Site-directed mutagenesis (SDM) protocol

For the amplification of *Ct*CBM6 site directed mutants, a pET-21a clone containing *Ct*CBM6 was used as a template (p*Ct*Xyn11A). Primers were constructed according to section 2.2.2.2 and are given in Table 2.5. The reaction for the amplification of the *Ct*CBM6 mutants is shown in Table 2.10 and the order the components appear in the table is the same as the order of addition to the PCR tube.

Tube component	Volume added	Final concentration in PCR reaction.
Sterile 18.2 MΩ H_2O	40.5 µl	
Forward primer 125 ng μl^{-1}	1 µl	2.5 ng μl^{-1}
Reverse primer 125 ng μl^{-1}	1 µl	2.5 ng μl^{-1}
Pfu reaction buffer 10 x	5 µl	1 x
p*Ct*Xyn11A (template) 500 ng μl^{-1}	1 µl	10 ng μl^{-1}
dNTPs 25 mM (Each)	0.5 µl	0.25 mM (Each)
Pfu polymerase (3 U μl^{-1})	1 µl	0.06 U μl^{-1}

Table 2.10: Reaction components for PCR of site-directed mutants.
For the recipe of *Pfu* reaction buffer (10 x), refer to Appendix A3. For primer sequences, refer to Table 2.5.

The Eppendorf MasterCycler™ machine was then programmed to proceed as shown in Table 2.11.

Temperature (°C)	Duration (min : sec)	Cycles
95.0	00:30	1
95.0 55.0 68.0	00:30 01:00 11:35	16
10.0	Continuous	

Table 2.11: Reaction conditions for PCR of site-directed mutants.

2.2.2.1.2 Polymerase chain reaction (PCR) protocol

For the amplification of HylP1, HylP2 & HylP3

Primers were constructed (Section 2.1.7.2) so as to anneal to the 5' and 3' ends of the sequence encoding the mature protein. Primers were also designed so as to provide recognition sites for restriction endonucleases upstream and downstream of the target sequence. These engineered sites allowed the open reading frame (ORF) to be inserted into a pET expression vector in the correct frame. Two PCR products of the same gene were amplified from the genomic DNA of *Streptococcus pyogenes* SF370 (ATCC 700294; Ferretti *et al.*, 2001). One difference between the two PCR products was that one product had a stop site in the reverse primer; hence the PCR products were named 'Stop' and 'No Stop'. PCR products generated from a reverse primer containing a stop site were used for subsequent cloning into pET-28a or pET-22b vectors so that the encoded products contained an N-terminal His_6 tag or no tag, respectively. PCR products

generated from a reverse primer containing no stop sites were used for subsequent cloning into pET-22b vector so that the encoded product contained a C-terminal His_6 tag. Also, each PCR product has a different annealing temperature for amplification. The gene sequences for HylP1 and HylP3 were very similar, thus it was necessary to amplify HylP1 and HylP3 from the generated larger fragment using primers that anneal to sequences either side of those genes that were dissimilar. Once these fragments had been cloned, a second round of PCR to amplify the specific genes was performed.

The reactions for PCR products 'Stop' and 'No Stop' were set up in sterile microcentrifuge tubes (0.2 ml) as detailed in Table 2.12. Platinum *Pfx* DNA polymerase contains a highly thermostable recombinant DNA polymerase from *Pyrococcus sp.* strain KOD and anti-*Pfx* DNA polymerase inhibitor antibodies (Westfall *et al.*, 1999). The antibodies keep the *Pfx* DNA polymerase inactive at ambient temperatures and thus reduces non-specific product formation. The inhibition of *Pfx* DNA polymerase is completely reversed at temperatures in excess of 65 °C, providing an automatic hot start that increases specificity, sensitivity, and yield, and allows room temperature assembly. The order the components appear in the table is the same as the order of addition to the PCR tube.

Tube component	Volume added	Final concentation in PCR reaction.
Sterile 18.2 MΩ H_2O	40.6 µl	
Forward primer 12.5 µM	1 µl	0.25 µM
Reverse primer 12.5 µM	1 µl	0.25 µM
Pfx reaction buffer 10 x	5 µl	1 x
S. pyogenes SF370 genomic DNA (ATCC 700294) 0.1 µg μl^{-1} or HylP1/HylP3 plasmid DNA	1 µl/0.5 µl	0.002 µg μl^{-1}
dNTPs 25 mM (Each)	0.4 µl	0.2 mM (Each)
Pfx polymerase (3 U μl^{-1})	1 µl	0.06 U μl^{-1}

Table 2.12: Reaction components for PCR.
Recipe for *Pfx* reaction buffer (10x) is patented. For primer sequences, refer to Table 2.6.

The tubes were then capped and centrifuged briefly to collect the contents. The Eppendorf MasterCycler™ machine was then programmed to proceed as shown

in Table 2.13. A gradient PCR protocol was used where the first five cycles of PCR were carried out at the primary annealing temperature and the subsequent 25 cycles of PCR were carried out at the secondary annealing temperature. The primary annealing temperature was calculated to be 5 °C lower than the lowest T_m of the primer pair (excluding the restriction site). The secondary annealing temperature was calculated to be 5 °C lower than the lowest T_m of the primer pair (including the restriction site). The primary and secondary annealing temperatures for the 'Stop' and 'No Stop' PCR products of each gene are given in Table 2.6.

Temperature (°C)	Duration (min: sec)	Cycles
94.0	02:00	First denaturation
94.0	00:15	5
1° annealing temperature	00:30	
68.0	02:00	
94.0	00:15	25
2° annealing temperature	00:30	
68.0	02:00	
68.0	10:00	Final extension
10.0	Continuous	

Table 2.13: Reaction conditions for PCR.
Refer to Table 2.6 for primary and secondary annealing temperatures for each gene to be amplified.

After the PCR reactions had cooled to 10 °C, a 10 µl aliquot was removed and analysed by agarose gel electrophoresis (Section 2.2.2.1.4) to check if the amplification had been successful.

2.2.2.1.3 Restriction digests

All restriction endonuclease (New England Biolabs) digests were performed according to the manufacturers instructions. Unless otherwise stated, each µg of DNA was cut with 5 Units of enzyme at 37 °C for > 1.5 h. Heat inactivation was carried out at 65 °C for 10 min when required. BSA was added to the reaction mix at a final concentration of 0.1 mg ml^{-1}, and the appropriate reaction buffer (Appendix A3) was also added to a final concentration of 1 x. Care was taken when preparing digests that the glycerol concentration did not exceed 10 % in the final reaction, in case enzyme specificity was affected.

Digests with two or more restriction enzymes were carried out simultaneously if the enzymes had high enough (> 50 %) activity in the same buffer. If the enzymes needed different buffers, then they were added sequentially. After ~ 1.5-2 h at 37 °C, the enzyme was then heat inactivated and the restriction digest was purified to remove the buffer using a NucleoSpin® Extract 2 in 1 kit (Section 2.2.2.1.7). The purified digest was then digested with the second restriction endonuclease with the appropriate buffer. The restriction endonucleases used to cut the plasmids/vectors are given in Table 2.14. A list of restriction endonucleases used in this study and corresponding reaction buffers are given in Appendix A3. The digested samples were then subjected to agarose gel electrophoresis (Section 2.2.2.1.4). The required DNA fragment was recovered from the gel using a NucleoSpin® Extract 2 in 1 kit and used for ligation (Section 2.2.2.1.7 and 2.2.2.1.12, respectively).

Plasmid/Vector	Restriction endonuclease
*pCt*Xyn11A	*Nde* I / *Xho* I
*p*HylP1: Stop	*Nde* I / *Bam*H I
No stop	*Nde* I / *Xho* I
*p*HylP2: Stop	*Nde* I / *Xho* I
No stop	*Nde* I / *Xho* I
*p*HylP3: Stop	*Nde* I / *Bam*H I
No stop	*Nde* I / *Xho* I
pET-21a	*Nde* I / *Xho* I
pET-22b	*Nde* I / *Bam*H I or *Nde* I / *Xho* I
pET-28a	*Nde* I / *Bam*H I or *Nde* I / *Xho* I

Table 2.14: Restriciton endonucleases used to cut plasmids/vectors.
For a list of corresponding reaction buffers, refer to Appendix A3.

2.2.2.1.4 Agarose gel electrophoresis

Agarose gels were prepared by heating agarose in 1 x TAE (Section 2.1.6.1.1) buffer until the solid had completely dissolved. The amount of agarose added depended on the size of the gel to be cast. All gels in this study were 1 % (w/v) agarose in a total volume of 30 ml, except those gel cast for purification purposes, which were made in a volume of 50 ml in order to give the wells a greater capacity.

Once dissolved, the solution was allowed to cool to ~ 60 °C before pouring the solution into a mini-gel casting tray (avoiding bubbles). The appropriate sized comb would then be inserted into the solution (for analytical purposes, a 12-

toothed comb able to hold 15 µl in each well was used, and for preparative purposes, a 5-toothed comb able to hold 50 µl in each well was used) and the gel allowed to set for > 20 min.

After the gel had set, the gel tray was placed horizontally in an agarose gel electrophoresis tank, submerged in 1 x TAE buffer (Section 2.1.6.1.1). The comb was then carefully removed and 10 µl samples were then prepared by the addition of 2 µl of sample buffer (Bromophenol blue 6 x; Section 2.1.6.1.2), and added to the wells. A 12 µl volume of an appropriate size standard (1 µg DNA; Section 2.1.6.1.3) would also be added to an unoccupied well in order to determine the size of electrophorised DNA fragments. The samples would then be electrophorised at 100 mA (200 V) for approximately 40 min to ensure good separation of DNA fragments.

2.2.2.1.5 Visualisation of DNA and photography of agarose gels

After electrophoresis, the agarose gel was removed from the apparatus and stained by submersion in a 10 µg ml^{-1} solution of ethidium bromide (Section 2.1.6.1.4) for 10 min with gentle shaking on a flat bed orbital shaker (~ 40 rpm). For analytical gels the ethidium bromide was re-used until staining became weak, however, preparative gels required fresh ethidium bromide each time. After staining, the gel was washed with distilled water and any DNA bands were visualised using a gel documentation system (Bio-Rad Gel Doc 2000 using Quantity One™ software). Hard copies of the gel picture were produced using a Mitsubishi Video Copy Processor (Model P91 attached to the gel doc system), with Mitsubishi thermal paper (K65HM-$_{CE}$ / High density type, 110 mm x 21 m).

2.2.2.1.6 Preparation of plasmid DNA (pDNA)

2.2.2.1.6.1 Standard preparation of pDNA

Purification of plasmid DNA from transformed *E. coli* strain XL1-Blue and XL10-Gold (Section 2.2.1.2.1) was achieved using a NucleoSpin® plasmid kit (Section 2.1.6.3). The method by which the purification was performed is detailed here.

A 3 ml volume of bacterial cell culture (Section 2.2.1.2.1) was pelleted in a 1.5 ml micro-centrifuge tube by repeated centrifugation at 14000 x g for 60 sec, and the supernatant discarded. The pellet was then pulsed and residual supernatant removed. The pellet was then re-suspended by vortexing in 250 µl of buffer A1 (cell re-suspension buffer). Buffer A2 (250 µl; cell lysis buffer) was then added and the sample was mixed by gentle inversion 5-6 times, prior to incubation at room temperature for 5 min. Buffer A3 (300 µl; precipitation of proteins and chromosomal DNA) was then added to the sample, and the sample was again mixed by gentle inversion 5-6 times, and incubated on ice for 5 min. Precipitated material was then separated by centrifugation at 14000 x g for 12 min, and the supernatant was retained and applied to a NucleoSpin® filter tube placed into a 2 ml receptacle centrifuge tube. The DNA was then loaded onto the filter by centrifugation at 14000 x g for 1 min and the flowthrough discarded, and washed with 500 µl of AW buffer, followed by an additional wash with 700 µl of buffer A4 (50 % ethanol), both washes were performed by centrifugation at 14000 x g for 1 min and in each case the flowthrough discarded. Prior to elution, the empty NucleoSpin® filter tube and receptacle were centrifuged at 14000 x g for 2 min in order to remove any residual ethanol. Finally, the NucleoSpin® filter tube was transferred into a sterile 1.5ml micro-centrifuge tube and the DNA was eluted from the filter by centrifugation at 14000 x g for 1 min after the addition of 50 µl of AE (elution buffer). The components of each buffer are given in Appendix A5.

2.2.2.1.6.2 Preparation of pDNA for sequencing

For the preparation of pDNA for sequence analysis, the standard preparation of pDNA method was employed (Section 2.2.2.1.6.1), but with some additions as detailed here. pDNA was washed twice with Buffer A4 (2 x 700 µl; 50 % ethanol) and eluted twice from the filter by the addition of 2 x 100 µl of sterile 18.2 MΩ H_2O pre-warmed to 70 °C in a water bath. The DNA was then concentrated by lyophilisation overnight in a freeze drier unit (Appendix D3) prior to reconstitution with TE buffer (Section 2.1.6.2.1)/18.2 MΩ H_2O to give a final concentration 0.2 µg μl^{-1}. This method of elution was found to give a higher yield of DNA, required for sequence analysis.

2.2.2.1.6.3 Crude preparation of pDNA for the screening of putative blue/white recombinants

A 3 ml volume of bacterial cell culture (Section 2.2.1.2.1) was pelleted in a 1.5 ml microcentrifuge tube by repeated centrifugation at 14000 x g for 1 min. The pellet was then pulsed and residual supernatant removed. The pellet was then re-suspended by vortexing in 150 µl of STETS buffer (Section 2.1.6.2.2) and 10 µl of 10 mg ml^{-1} hen egg white lysozyme (Section 2.1.6.4) was added to lyse the cells. Samples were then mixed by gentle inversion 5-6 times and incubated at room temperature for 15 min. The sample was then boiled for 40 secs and centrifuged at 14000 x g for 15 min immediately after removal from the water bath. The viscous pellet was then removed and discarded. To the remaining supernatant, 150 µl of chilled isopropanol was added and the sample mixed, prior to incubation at –20 °C for 1 h to allow precipitation of DNA. Upon removal from –20 °C the sample was immediately centrifuged at 14000 x g for 5 min and the supernatant discarded. The pellet was then pulsed and residual supernatant removed. The pellet was then washed with 500 µl of ethanol prior to centrifugation at 14000 x g for 3 min. After discarding the supernatant the pellet was then pulsed and the residual supernatant removed. Finally, the sample was desiccated for 10 min in order to remove any latent ethanol, prior to re-suspension of the pellet in 30 µl of TE buffer (Section 2.1.6.2.1) containing 100 µg ml^{-1} RNase (final concentration; Section 2.1.6.4).

2.2.2.1.7 Agarose gel purification

Purification of DNA fragments from agarose gels was achieved using a NucleoSpin® Extract 2 in 1 kit (Section 2.1.6.3). The method by which the purification was performed is detailed here.

Prior to removal of the fragment, the gel was 'briefly' visualised under low intensity UV and etching using a clean scalpel blade indicated the location of the fragment to be excised. The section of gel containing the fragment was then excised using a scalpel, and inserted directly into a pre-weighed 1.5 ml micro-centrifuge tube. The micro-centrifuge tube was then re-weighed and the weight of the gel slice it contained was calculated. For every 100 mg of agarose gel, 300 µl of buffer NT1 was added to the tube. The gel was then dissolved by incubation at

50 °C for 10 min with a brief vortex every 2-3 min. Once dissolved, the sample was then loaded into a NucleoSpin® filter tube and placed into a 2 ml receptacle centrifuge tube. The sample was then centrifuged at 6000 x g for 60 sec, and the flowthrough discarded. The sample was then washed by the addition of 700 µl buffer NT3, which was centrifuged at 14000 x g for 60 sec and the flowthrough discarded. The wash step was then repeated, prior to centrifugation at 14000 x g for 60 sec of the empty NucleoSpin® filter tube and receptacle, in order to remove any residual ethanol from the filter. The NucleoSpin® filter tube was then transferred into a sterile 1.5 ml micro-centrifuge tube and 30 µl of NE buffer was added precisely to the top of the filter membrane. The sample was then centrifuged at 14000 x g for 60 sec. The filter was then discarded and the DNA that was eluted into the micro-centrifuge tube was retained for subsequent use. To check the purification had worked, 2 µl was removed from the eluted sample, and made up to 10 µl with 18.2 MΩ H_2O. This sample was then analysed by agarose gel electrophoresis as detailed in section 2.2.2.1.4. The components of each buffer are given in Appendix A5.

2.2.2.1.8 PCR / RE digest clean up

Purification of double stranded DNA (dsDNA) from the other components of a PCR / RE digest reaction was achieved using a NucleoSpin® Extract 2 in 1 kit (Section 2.1.6.3). The method by which the purification was performed is detailed here.

Firstly, the volume of the reaction mix was adjusted to 100 µl using TE buffer (Section 2.1.6.2.1). Buffer NT2 (400 µl) was then added and the sample mixed. The sample was then loaded into a NucleoSpin® filter tube and placed into a 2 ml receptacle centrifuge tube. The sample was centrifuged at 6000 x g for 60 sec, and the flowthrough discarded. The sample was then washed by the addition of 700 µl of buffer NT3, which was centrifuged at 14000 x g for 60 sec and the flowthrough discarded. The wash step was then repeated; prior to centrifugation at 14000 x g for 60 sec of the empty NucleoSpin® filter tube and receptacle, in order to remove any residual ethanol from the filter. The NucleoSpin filter tube was then transferred into a sterile 1.5 ml micro-centrifuge tube and 30 µl of NE buffer was added precisely to the top of the filter membrane. The sample was then

centrifuged at 14000 x g for 60 sec. The filter was then discarded and the eluted DNA in the micro-centrifuge tube retained for subsequent use. To check the purification had worked 2 µl was removed from the eluted sample, and made up to 10 µl with 18.2 MΩ H_2O. This sample was then run on an agarose gel as detailed in section 2.2.2.1.4.

2.2.2.1.9 Spectrophotometric quantification of DNA

DNA concentration was quantified accurately by scanning through 200-300 nm using a UV spectrophotometer.

DNA was quantitatively determined by preparing a dilution of DNA in a 500 µl total volume. The dilution was then transferred to a 500 µl quartz cuvette and scanned in a UV vis spectrophotometer (previously zeroed with 18.2 MΩ H_2O) between the wavelengths of 200 and 300 nm (Figure 2.1). Optimal wavelength for the absorbance of DNA was assumed to be 260 nm, and 280 nm was assumed to be the optimal wavelength for contaminating protein. dsDNA at 50 µg ml^{-1} gives an A_{260} reading of 1.0, as does 33 µg ml^{-1} of single-stranded DNA (ssDNA) or oilgonucleotide. The concentration and purity of the DNA was then determined as detailed below.

[dsDNA] µg $µl^{-1}$ = (A_{260}) x 50 x dilution / 1000.
DNA purity ratio = $(A_{260})/(A_{280})$

A DNA purity ratio of 1.5-1.8 was accepted as minimal protein contamination.

Determination of the concentration of ssDNA was conducted in exactly the same way as for plasmid DNA except the calculation was altered as below.

[ssDNA] µg $µl^{-1}$ = (A_{260}) x 33 x dilution / 1000.

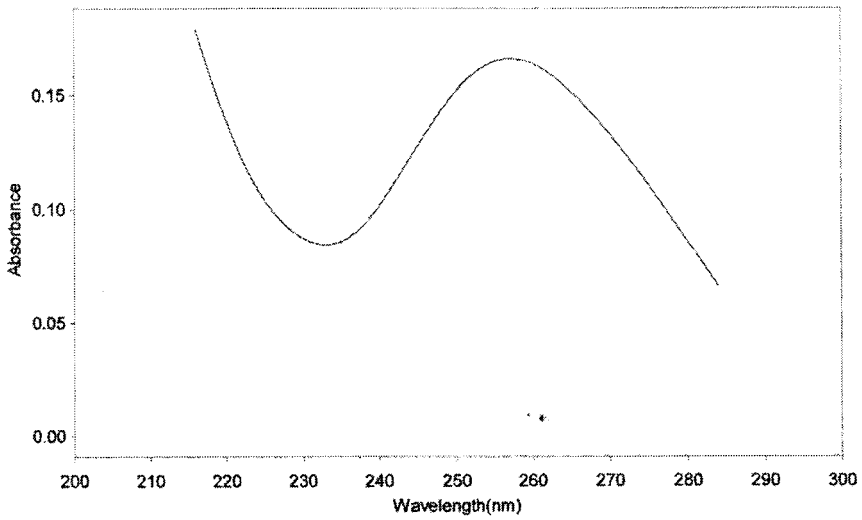


Figure 2.1: Spectrophotometric scan of DNA between 200 and 300 nm.
A UV visible Heλios α Spectronic Unicam spectrophotometer was used to obtain scan data.

2.2.2.1.10 Automated DNA sequencing

DNA (2 μg) was produced (Section 2.2.2.1.1) and diluted with 18.2 MΩ H_2O to a final concentration of 0.2 μg μl^{-1}. The DNA was sent to the Facility for Molecular Biology, Newcastle University, and was sequenced on ABI Prism™ Ready Reaction DyeDeoxy™ Terminator Cycle Sequencing apparatus, and analysed using an Applied Biosystems 373 A automated sequencer, to determine whether the single codon mutations had been created successfully. The standard sequencing primers T7 promoter or T7 terminator were used.

2.2.2.1.11 A - tailing DNA

In preparation for ligation into the pGEM®-T Easy vector (Section 2.1.8), A-overhangs were synthesised on the DNA fragment using the method described below.

7 μl of purified PCR product (Section 2.2.2.1.8) was made up to a 10 μl reaction volume by addition of the components listed in Table 2.15.

Component	Volume	Final concentration in A -tailing reaction
Purified PCR product	7 µl	
Taq DNA polymerase reaction buffer (10 x)	1 µl	1 x
$MgCl_2$ (25 mM)	0.5 µl	1.25 mM
dATP (4 mM)	0.5 µl	0.2 mM
Taq DNA polymerase 5 U μl^{-1}	1 µl	0.5 U μl^{-1}

Table 2.15: Reaction components of the A –tailing procedure.
The composition of *Taq* DNA polymerase reaction buffer is given in Appendix A3.

Once the reaction mix was complete, the tube was given a pulse in a micro-centrifuge, prior to incubation in a 70 °C water bath for 30 min. The DNA was then used for ligation into pGEM®-T Easy vector (Section 2.2.2.1.12).

2.2.2.1.12 Ligation of DNA

For most ligations, a 2:1 insert to vector ratio was required. Concentrations were determined by running samples on an agarose gel and using the analysis tool of the Bio-Rad Gel Doc 2000 Quantity One™ software.

2.2.2.1.12.1Ligation into pGEM®-T Easy vector

Ligation of cohesive termini, generated from A-tailing blunt PCR products (Section 2.2.2.1.11), into linearised pGEM®-T Easy vector (Table 2.2) was performed by addition of the components listed in Table 2.16.

Component	Volume
pGEM®-T Easy vector (de-phosphorylated & linearised; 50 ng; Lab stock)	1 µl
2 x Rapid ligation buffer	5 µl
A- tailed PCR product (33 ng; Section 2.2.2.1.11)	3 µl
T_4 DNA Ligase (3 U μl^{-1})	1 µl

Table 2.16: Reaction components for ligation into pGEM®-T Easy vector.
The composition of Rapid ligation buffer is given in Table 2.4.

After addition of reaction components, the ligation was incubated at room temperature for 1 h, or at 4 °C overnight prior to transformation into XL1-Blue cells (Section 2.2.2.1.13).

2.2.2.1.12.2 Ligation into pCR®-Blunt vector

Ligation of blunt-ended PCR products (Section 2.2.2.1.2) into linearised pCR®-Blunt (Table 2.2) was performed by the addition of the components listed in Table 2.17. A 10:1 molar insert to vector ratio was used.

Component	Volume
pCR®-Blunt vector (linearised; 25 ng)	1 µl
10 x Ligation buffer (with ATP)	1 µl
Blunt PCR product (50 ng; Section 2.2.2.1.2)	5 µl
18.2 MΩ H_2O	2 µl
T_4 DNA Ligase (4 U μl^{-1})	1 µl

Table 2.17: Reaction components for ligation into pCR®-Blunt vector.

This reaction was then incubated at 16 °C overnight prior to transformation into TOP10 chemically competent cells (Table 2.2; Invitrogen Corp.).

2.2.2.1.12.3Ligation into pET vector

Ligation into pET vector (Table 2.2) was performed by the addition of the components listed in Table 2.18.

Component	Volume
pET vector cut with appropriate REs (50 ng; de-phosphorylated; Lab stock)	2 µl
10 x Ligation buffer	1 µl
Fragment cut from pGEM®-T/pCR®-Blunt vector and gel purified (100 ng; Section 2.2.2.1.7)	2 µl
18.2 MΩ H_2O	4 µl
T_4 DNA Ligase(3 U μl^{-1})	1 µl

Table 2.18: Reaction components for ligation into pET vector.

After addition of reaction components, the ligation was incubated at 4 °C overnight prior to transformation into XL10-Gold cells (section 2.2.2.1.13).

2.2.2.1.13 Chemical transformation

2.2.2.1.13.1 Transformation using *E. coli* XL1-Blue, XL10-Gold, BL21 (DE3) and B834 (DE3) cells

Transformations were performed using prepared chemically competent cells (Section 2.2.1.4) in 50 µl aliquots. Firstly, the DNA to be transformed was added and mixed gently with a pipette tip and incubated with the cells on ice for 20 min. The volume of DNA added from a ligation was 2 µl, and from a plasmid preparation was 1 µl. The cells were then transferred to a water bath at 42 °C and

heat shocked for 1 min, prior to a second incubation on ice for 2 min. The cells were then recovered by the aseptic addition of 450 µl of room temperature NZY^+ enrichment broth (Section 2.1.3.1.2), and incubation at 37 °C for 1 hour shaking (horizontally for maximum aeration) at 220 rpm.

Recovered cells were then aseptically plated on media appropriate to the vector transformed (Section 2.2.1.3). Bacteria transformed with pET plasmids were plated on the appropriate antibiotic. Plates were then incubated inverted for ~ 16 h at 37 °C to allow the growth of bacterial colonies.

2.2.2.1.13.2 Transformation using One shot™ TOP10 cells

One shot™ TOP10 cells (25 µl; Table 2.2) were thawed on ice for each ligation/transformation. The pCR®-Blunt ligation (2 µl) was added to the cells and mixed gently with a pipette tip. The cells/ligation mix was then incubated on ice for 30 min. The cells were transferred to a 42 °C water bath and heat-shocked for exactly 45 sec. Pre-warmed (37 °C) SOC medium (125 µl) provided with the One shot™ TOP10 cells was added and then incubated at 37 °C for 1 h with orbital shaking (horizontally for maximum aeration) at 220 rpm.

Each transformation was then plated (Section 2.2.1.3) on LB-kanamycin plates and incubated inverted for ~ 16 h at 37 °C to allow the growth of bacterial colonies.

2.2.2.2 Construction of mutant plasmids of *Ct*CBM6 by site directed mutagenisis (SDM)

SDM was performed according to the Stratagene instruction manual for the QuikChange™ Site-Directed Mutagenesis Kit. However, the components of the protocol were obtained from sources other than Stratagene.

Firstly, mutagenic oligonucleotide primers (Section 2.1.7.1) with a length of 25-45 bases, and GC content ≥ 40 °C were designed with the mutation in the centre of the primer and a T_m (melting temperature) of ≥ 78 °C. The T_m for each primer was determined as follows:

$$T_m = 81.5 + 0.41\ (\%\ GC) - 675 / N - \%\ mismatch$$

Key: % GC = Percentage of guanine and cytosine residues in the primer
N = Primer length in bases
% mismatch = Percentage of the primer non complementary at the point of mutation

Mutant plasmid was then generated by PCR (Section 2.2.2.1.1). The PCR template DNA (i.e. the non-mutated supercoiled DNA p*Ct*Xyn11A) was then digested in a waterbath at 37 °C for 1 h by adding 1 µl of *Dpn* I (10 U μl^{-1}) restriction enzyme (Section 2.1.6.4) directly to the cooled PCR reaction and mixed gently. The *Dpn* I treated DNA was then transformed into *E. coli* XL10-Gold (Section 2.2.2.1.13.1), and plasmid DNA (Section 2.2.2.1.6.2) made from one of the transformants was sent to the Central Molecular Biology Facility at the University of Newcastle-Upon-Tyne for sequence analysis in order to confirm the mutation (Section 2.2.2.1.10).

2.2.3 Protein methods

2.2.3.1 SDS - PAGE electrophoresis

SDS – PAGE was conducted according to the method of Laemmli (1970).

SDS - PAGE gels were cast between two clean glass plates of dimensions (10.1 x 7.2 and 10.1 x 8.2 cm) separated by a 1 mm spacer ridge on the larger of the two plates. The plates were clamped together and checked to ensure the bottoms were flush, prior to securing the gel with vertical downward pressure in a casting stand on a rubber gasket. The components for the resolving gel (Section 2.1.9.1.1) were then added together, mixed, and then pipetted into the space between the plates, until the solution was approximately 2 cm below the top of the smallest plate. In order to prevent oxygen from inhibiting polymerisation and to produce a straight top edge, a small volume of water was then added to the top of the solution. The resolving gel was then allowed to polymerise for ~ 45 min. Once set, the water on top of the gel was removed, and the components of the stacking gel were added (Section 2.1.9.1), mixed, and pipetted up to the top of small plate. A 10-toothed

comb (1.1 x 0.75 cm) was then immediately inserted into the gap between the plates. The stacking gel was then allowed to polymerise for ~ 30 min prior to removal of the comb to reveal the wells, which were immediately rinsed with water to remove any gel debris. The gel was then placed vertically into the buffer tank (Appendix D8) secured in a holder with another gel constructed in exactly the same manner to the opposite side. SDS-PAGE (1 x) running buffer (Section 2.1.9.1.3) was then poured into the tank ensuring the level of buffer in the space between the two gels was higher than the level on the outside of the gels.

Samples were then prepared by the addition of 5 µl of SDS-PAGE loading buffer (Section 2.1.9.1.4) to 20 µl of sample, which was then pulsed briefly prior to boiling for 3 min. Size standards (Section 2.1.9.1.6) were prepared in the same way, except 8 µl of loading dye was added to 4 µl of size standard. All samples and standards were then centrifuged at 14000 x g for 1 min prior to loading 20 µl of all samples and 10 µl of all standards into the wells using a Hamilton syringe. The pair of gels were then electrophorised at 120 milliamps, 200 volts for ~ 50 mins, or until the bromophenol blue dye had dropped off the bottom of the gel.

2.2.3.1.1 Visualisation of protein bands and photography of SDS-PAGE gels

After electrophoresis, the gels were removed from the apparatus and stained for 10 min in Coomassie blue stain (Section 2.1.9.1.7), followed by immersion in destain (Section 2.1.9.1.8) until crisp band images were visible against a clear background. The gels were then photographed using a gel documentation system (Bio-Rad Gel Doc 2000 using Quantity One™ software). Hard copies of the gel picture were produced using a Mitsubishi Video Copy Processor (Model P91 attached to the gel doc system), with Mitsubishi thermal paper (K65HM-$_{CE}$ / High density type, 110 mm x 21 m).

2.2.3.2 Isolation of periplasmic fraction (PPF) from bacterial cells

Proteins that were exported to the *E. coli* periplasm were obtained by the osmotic shock method of Hsuing *et al.* (1986). Induced cells were concentrated by centrifugation for 10 min at 4500 x g at 4°C, after which the supernatant was decanted and the cell pellet was gently re-suspended in 10 ml ice-cold 50 mM sodium phosphate buffer, pH 7.0 containing 20 % (w/v) sucrose (Section

2.1.9.4.1). The cells were then incubated on ice for 30 min, followed by centrifugation for 10 min at 4500 x g at 4 °C. The supernatant was discarded and the pellet was gently resuspended in 10 ml ice-cold 18.2 MΩ H_2O and incubated on ice for a further 30 min, before being centrifuged for 20 min at 10 000 x g at 4°C. The supernatant was carefully decanted into dialysis tubing (pre-boiled for 10 min in 18.2 MΩ H_2O; section 2.1.9.5.1), which was sealed with two knots at one end and closures at the other, and the proteins dialysed twice against 8 L 50 mM sodium phosphate buffer, pH 7.0 overnight at 12 °C to remove sucrose (Section 2.1.9.5), before being aliquotted into 1.0 ml fractions and stored at – 20 °C until use.

2.2.3.3 Isolation of cell free extract (CFE)

To obtain CFE, firstly induced cells were concentrated by centrifugation for 10 min at 5500 x g, after which the supernatant was decanted and the cells were re-suspended in 5-20 ml of cell re-suspension buffer (Section 2.1.9.4.2). Cells were then lysed in 5 ml volumes by sonication on ice at amplitude 14 for 1 min in 10 sec bursts with 10 sec intervals. The lysate was then decanted into centrifuge tubes and centrifuged at 4 °C for 20 min at 24 000 x g. After centrifugation the soluble fraction was decanted into a plastic universal and maintained on ice as CFE.

2.2.3.4 Protein purification

Recombinant proteins used in this study were purified by a number of different methods. In all cases, the purity of the protein was assessed by SDS-PAGE (Section 2.2.3.1). Immobilised metal affinity chromatography (IMAC) was used to purify recombinant proteins, under native conditions, which contained a stretch of 6 Histidine residues at the N-terminus (His_6 tag). The metal affinity matrix used was Talon™ or Sepharose™ chelating fast flow resin (Section 2.1.9.4.3).

2.2.3.4.1 Preparation of Ni^{2+}/ Co^{2+} affinity resin

The resin was supplied as 75 % slurry in 20 % ethanol. This procedure results in a bed volume of 3 ml charged with Ni^{2+} (Sepharose™ chelating fast flow resin; section 2.1.9.4.3) or Co^{2+} (Talon™ resin; section 2.1.9.4.3) ions.

Firstly, the resin (Sepharose™ resin only; Talon™ is purchased ready charged) was mixed and 3.66 ml was pipetted into a sterile universal (to give a bed volume of 3 ml). The resin was sedimented by centrifugation at 500 x g for 2 min and the supernatant discarded. The resin was then washed by the addition of 10 ml of 18.2 MΩ H_2O and mixed by end-over-end rotation for 5 min. The resin was then re-sedimented by centrifugation at 500 x g for 2 min and the supernatant again discarded. The resin was then charged by the addition of 1 ml of $NiSO_4$ (1 M) and mixed for 5 min by end-over-end rotation, prior to centrifugation at 500 x g for 2 min and removal of the supernatant to waste. Again the resin was washed by the addition of 10 ml 18.2 MΩ H_2O, and mixed by end-over-end rotation prior to re-sedimentation by centrifugation at 500 x g for 2 min. This wash step was repeated twice more, prior to the addition of one gel volume of start buffer (Section 2.1.9.4.5).

For use in a stepped elution (Section 2.2.3.4.2), the prepared chelating Sepharose™ resin was then decanted into a pre-prepared PD-10 column with a pre-dampened filter at the bottom of the column and the cap in place. Talon™ resin (3.66 ml) was added to a prepared PD-10 column prior to the addition of one gel volume of wash buffer (Section 2.1.9.4.4).

Alternatively, the above procedure was scaled up to accommodate a 50 ml bed volume and the chelating Sepharose™ resin was poured into a C series column (C 16/20; Amersham Biosciences) of the dimensions 1.6 cm internal diameter (i.d.) x 20 cm length (L), attached to a FPLC® system and peristaltic pump. This method allowed the large scale purification of recombinant proteins using a gradient elution (Section 2.2.3.4.2).

2.2.3.4.2 Affinity purification using a fast flow Ni^{2+}/ Co^{2+} column

Recombinant proteins fused to a His_6 tag were purified under native conditions. For the purification of protein using the IMAC technique, the column was prepared (Section 2.2.3.4.1) and equilibrated by passing though 40 ml of wash buffer (Talon™, section 2.1.9.4.4) or start buffer (Chelating Sepharose™; section 2.1.9.4.5). PPF (Section 2.2.3.2) was prepared from 1 x 400 ml culture in 50 mM sodium phosphate buffer, pH 7.0, and applied to a 3 ml bed volume Talon™

column (for wild type *Ct*CBM6) or 3 ml bed volume of Sepharose™ resin (for *Ct*CBM6 site-directed mutants), and was incubated for 5 min at room temperature with end-over-end rotation. The column was subsequently washed with 30 ml of wash buffer (Section 2.1.9.4.4 and 2.1.9.4.5, respectively) to remove non-specific protein. Finally, the protein was eluted from the column by the addition of 2 x 5 ml of elution buffer (Section 2.1.9.4.4 and 2.1.9.4.5, respectively) and collected on ice in a sterile universal.

For the high yield purification of proteins (HylP1, HylP2 and HylP3) using the automated FPLC® system (Appendix D11 for automotive parts) gradient elution technique, the column was prepared by adding ~ 50 ml Sepharose™ chelating fast flow resin (Section 2.1.9.4.3) to a C series column (C 16/20; Amersham Biosciences). The column adapters were securely attached and the column was connected to the FPLC® system. Once connected, 1 M $NiSO_4$ solution was passed through the resin using the peristaltic pump until all the resin was a uniform colour. The column was then washed by passing through ~ 200 ml of Solution A (Section 2.1.9.4.6) using the FPLC® machine. Once the column had been equilibrated, CFE (10 –20 ml) was then loaded onto the column using a peristaltic pump. The column was then washed with 100 ml of Solution A (Section 2.1.9.4.6) at a volume flow rate of 5 ml/min (linear flow rate 149 cm/h) generated by the FPLC gradient pump. Finally, the protein was eluted over 50 min at a volume flow rate of 5 ml/min (linear flow rate 149 cm/h) using a linear gradient of imidazole (Solution B; section 2.1.9.4.6) extending from 10 mM to 500 mM, using Solution A (Section 2.1.9.4.6) as a diluent. The position of the target protein was determined by monitoring A_{280} using an inline UV spectrometer connected to the FPLC, and samples were collected in 5 ml volumes on a fraction collector. For recombinant proteins expressed in *E. coli* B834 (DE3), 10 mM DTT (final concentration) was added to the tubes in the fraction collector prior to elution of the protein. To confirm purity a 20 µl aliquot from each 5 ml fraction thought to contain the target protein was analysed by SDS - PAGE (Section 2.2.3.1) and the pure fractions pooled.

2.2.3.5 Gel filtration chromatography

Purified protein from an affinity chromatography column (Section 2.2.3.4.2) was exchanged by concentration (Section 2.2.3.6) into gel filtration buffer (Section 2.1.9.4.7). The gel filtration column (Section 2.1.9.4.3), of dimensions 1.6 cm (i.d.) x 60 cm (L), with a bed volume of 120 ml was prepared by equilibration (using the FPLC® system) at a volume flow rate of 1 ml/min with 120 ml of gel filtration buffer (Section 2.1.9.4.7). Protein sample, 0.5 ml, was then loaded onto the column and eluted over 120 min with gel filtration buffer running at a volume flow rate of 1 ml/min. The elution point of the target protein was identified by A_{280} using the inline UV spectrophotometer, and samples were collected on a fraction collector in 4 ml volumes. To confirm purity, a 20 µl aliquot from each 4 ml fraction thought to contain the target protein was analysed on a SDS-PAGE gel (Section 2.2.3.1) and the pure fractions were pooled.

2.2.3.6 Concentration and buffer exchange of protein

Concentration of protein was performed at 4 °C by centrifugation at 4000 x g in 10 kDa Vivaspin 6ml cut-off concentrator units (Viva Science). Protein was also buffer exchanged by dilution within concentrator units.

Alternatively, buffer exchange was performed by hydrating 2 g of Sephadex G-25 resin with the buffer to be exchanged into (Section 2.1.9.4.8.2). The resin was allowed to hydrate for 10 min prior to pouring into a PD-10 column with a pre-damped filter at the bottom. Excess buffer was allowed to flow through the column and the resin allowed to sediment prior to addition of a pre-damped filter to the top of the resin. A 40 ml volume of buffer (Section 2.1.9.4.8.2) was then washed through the column, and then 2.5 ml of protein (Section 2.2.3.4.2) was added, in the buffer to be removed. Finally, the protein was eluted from the column by the addition of 3.5 ml of buffer (Section 2.1.9.4.8.2). The column was then washed with 8 ml of buffer (Section 2.1.9.4.8.2) and was able to be used again for the same application.

2.2.3.6.1 Buffer exchange of *Ct*CBM6 using dialysis

Buffer exchange was performed by adding the purified protein to prepared dialysis tubing (Section 2.1.9.5.1) and securing tightly with closures at either end. The eluate was then dialysed overnight in the cold room (~ 12°C) in 8 L of 20mM sodium acetate buffer pH 5.0 (for crystallisation) or 50mM sodium phosphate buffer pH 7.0 (for binding studies). The dialysis was repeated overnight in 8 L of fresh buffer. The dialysed protein was then pooled into a sterile universal and stored on ice for future use. A 20 µl aliquot of the dialysed protein was analysed by SDS-PAGE (Section 2.2.3.1) to check purity.

2.2.3.7 Determination of protein concentration

2.2.3.7.1 Spectrophotometric method

Pure protein concentration was determined by the method of Gill and von Hippel (1989) and Pace *et al.* (1995). *Ct*CBM6 protein was suitably diluted in the appropriate buffer and the A_{280} determined by scanning from 240-340 nm (Figure 2.2). The molar absorptivity (ε) was calculated from the following equation:

$$\varepsilon\ (280)\ (M^{-1}\ cm^{-1}) = (\#\ Trp)\ (5700) + (\#\ Tyr)\ (1300)$$
$$= (1)\ (5700) + (6)\ (1300)$$
$$\varepsilon = 13\ 500\ M^{-1}\ cm^{-1}$$

Protein was in 50 mM sodium phosphate buffer, pH 7.0. The scan was performed in a 1 ml path length quartz cuvette at 2500 nm/min scan speed. A blank of buffer only was used.

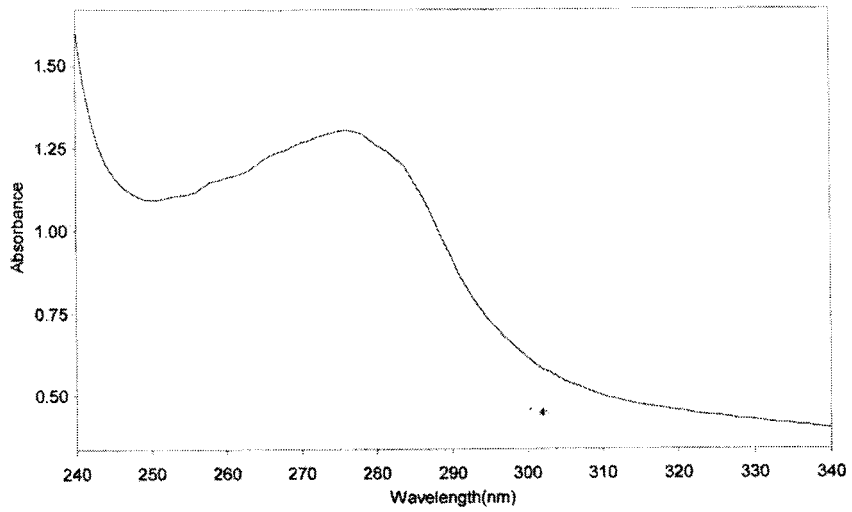


Figure 2.2: Spectrophotometric scan of pure *Ct*CBM6 between 240 and 340 nm. A UV visible Heλios α Spectronic Unicam spectrophotometer was used to obtain scan data.

Once the A_{280} and molar absorbtivity had been obtained, the concentration of *Ct*CBM6 could be determined, in μM, using the following equation:

$$(10^6 / \varepsilon)\ (A_{280})\ (\text{Dilution factor}) = \text{Protein concentration}\ (\mu M)$$

2.2.3.7.2 Bradford method

The amount of protein present in samples was determined by the method of Bradford (1976).

Protein determination by the Bradford method involved the construction of 10 dilutions of BSA in 500 μl aliquots ranging from 1-10 μg ml^{-1}. Bradford's solution (500 μl; Section 2.1.9.8) was then added to each dilution, and mixed. Each standard was then measured in a glass cuvette at A_{595} against a blank containing 500 μl of 18.2 MΩH_2O mixed with 500 μl of Bradford's solution. The glass cuvette was washed with methanol after every measurement, and all measurements were completed within 40 min of initially adding the Bradford's solution. A standard curve was then constructed (Figure 2.3). Dilutions of the protein to be determined were then made in 500 μl volumes and measured after

the addition of Bradford's solution in the same way as the standards. Those values that fell on the standard curve were then used to calculate protein concentration, using the equation below.

Protein conc mg ml^{-1} = (Conc obtained from the curve) x (dilution of sample) / 1000

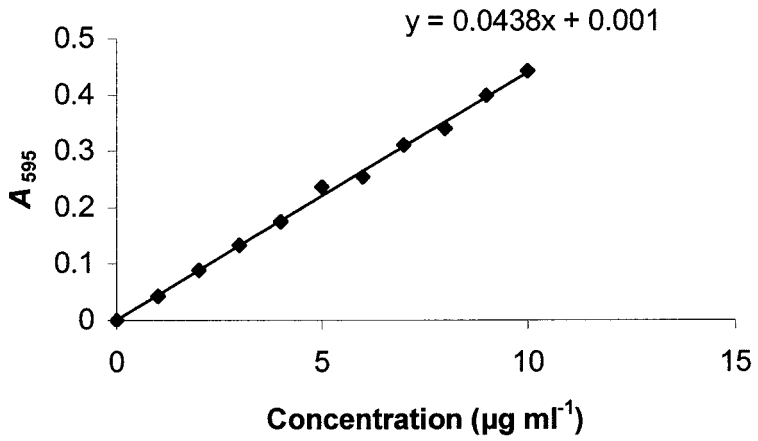


Figure 2.3: Bradford assay standard curve for the quantification of protein.
The curve was generated by the addition of 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, + 10 µg ml^{-1} BSA in 500 µl 18.2 MΩH_2O to 500 µl Bradford's reagent. After 5 min at room temperature, the A_{595} of the solution was determined.

2.2.3.8 Crystallisation

2.2.3.8.1 Silianisation of coverslips for crystallisation

The coverslips (Menzel-Glaser, 18 x 10 mm) for crystallisation were thoroughly cleaned with detergent and rinsed thoroughly after washing. A 1 % (v/v) working solution of Aqua Sil (Hampton Research) was prepared. The cleaned coverslips were then immersed in the Aqua Sil working solution one-by-one, stirred briefly with a glass rod, and left for 30 min. The coverslips were then washed with tap water before washing in 18.2 MΩ H_2O and finally dried thoroughly before use.

2.2.3.8.2 The hanging drop method

Crystals of HylP1, HylP2, and HylP3 were grown by vapour phase diffusion using the hanging drop method.

Crystal screening was performed in 24-multi well plates (Falcon® Multiwell™ 24 well; Becton Dickinson), into which 1.0 ml of crystallisation solution (Appendix C) was aliquoted into each well. Wells were then greased around the rim using vacuum grease (Dow Corning®), dispensed from a syringe. Cover slips were silanised by pre-treatment with Aqua Sil solution (Section 2.2.3.8.1), and polished with a silk scarf prior to the addition of protein drops.

Purified protein was then exchanged into 10 mM HEPES, 50 mM $CaCl_2$, pH 7.4, by the concentration method (Section 2.2.3.6) and concentrated to between 10 – 20 mg ml^{-1}. In some cases, a 1 M buffer (Sodium acetate, pH 4.4; HEPES, pH 7.25; Tris pH 9.0) was added at 10 % (v/v) to adjust the pH of the protein solution.

Finally, 1 µl of protein was added to 1 µl of mother liquor (removed from the well) on a cover slip to form a 2 µl drop. The cover slip was then inverted and sealed above the appropriate well. Once this procedure had been repeated for all crystallisation solutions (Appendix C), the plates were then marked appropriately and incubated at 22 °C.

The conditions of any crystals that grew in the crystallisation screens were then used for further optimisation studies. This involved varying the concentration of protein, salt, pH, and addition of cryoprotectants (e.g. MPD (2-methyl-2,4-pentanediol), glycerol, ethylene glycol; 10 – 30 %) to either the mother liquor or the drop in order to obtain crystals that would easily withstand the cryo-protectant prior to freezing more. Once native crystals had grown, Se-Met-substituted protein for HylP1, HylP2, and HylP3 were made (Section 2.2.1.2.4) and crystallisation studies using the same native crystallisation conditions were set up, except that the protein additionally contained 1 mM DTT (not in mother liquor).

2.2.3.8.3 The Seeding Method

The technique of seeding crystals was used in order to try and encourage the Se-Met-substituted HylP1 protein to grow. Crystal microseeding was performed in 24-multi well plates, into which 1.0 ml of 3.25 M sodium formate solution was added to each well. The wells were then greased round the rim using vacuum

grease dispensed from a syringe. Native HylP1 protein (1 µl) was then added to 1 µl of mother liquor from the well and 1 µl 15 % MPD solution on a coverslip, and this process was repeated another five times. Once the drops had been set up, a native crystal of HylP1, from previous crystallisation studies, was crushed to produce microcrystals, using a needle, and a human hair was then run through these microseeds. The hair, coated in microseeds, was run through the drops set up above and the coverslips inverted and sealed above the appropriate well. This process was then repeated for the Se-Met-substituted HylP1 protein. The plate was then stored at 22 °C.

2.2.3.8.4 The provision of heavy metal derivatives

The $Ta_6Br_{12}^{2+}$ cluster compound is known to be a powerful reagent for derivatization of crystals of large macromolecules at low resolution (Banumathi *et al.*, 2003). The cluster is a regular octahedron of six tantalum atoms with twelve bridging bromine atoms at the edges of the octahedron. Single hexagonal crystals of HylP1, previously grown by vapour phase diffusion using the hanging-drop method (Section 2.2.3.8.2), were transferred using a rayon-fibre loop directly into a well containing 0.5 ml of mother liquor (3.25 M sodium formate). Small specs of the Ta_6Br_{12} compound were then placed around the edges of the crystal using a hypodermic needle. The well was then covered and left at 22 °C until the Ta_6Br_{12} compound had diffused into the crystal. Once the crystals were a homogeneous green colour, the crystals were frozen in liquid nitrogen, using a rayon-fibre loop (Section 2.2.3.8.5), before being transferred into a liquid nitrogen cold stream ready for testing for diffraction quality (Sections 2.2.3.8.6 and 2.2.3.8.7).

Heavy metal derivatives were also obtained by soaking native HylP1 crystals, previously grown by vapour phase diffusion using the hanging-drop method (Section 2.2.3.8.2), in platinum (II) 2,2'6'2-terpyridine chloride dihydrate, uranyl acetate and methyl mercury (II) chloride. A stock solution (50 mM) of each of the heavy metal salts was prepared before diluting to a final concentration of 5 mM in the mother liquor from the original screen well (3.25 M sodium formate). HylP1 crystals were soaked in platinum (II) chloride, uranyl chloride and methyl mercury (II) chloride for 30 min, overnight, and 90 min, respectively. Crystals were then frozen in liquid nitrogen, using a rayon-fibre loop (Section 2.2.3.8.5),

before being transferred into a liquid nitrogen cold stream ready for testing for diffraction quality (Sections 2.2.3.8.6 and 2.2.3.8.7).

2.2.3.8.5 Mounting crystals and optimisation of cryoprotectant conditions

MPD and glycerol were increasingly added to the mother liquor at intervals of 5% (v/v). Samples were frozen in a rayon-fibre loop and subjected to X-ray analysis. The point at which ice rings were not visible in the diffraction pattern indicated a suitable additive concentration to act as a cryoprotectant (25% MPD for HylP1 and 25% for glycerol HylP3). The mother liquor solutions were re-made with the optimised MPD/glycerol concentration and these served as the cryoprotectant solutions. For the heavy metal derivatives of platinum (II) chloride, uranyl acetate and methyl mercury (II) chloride, a 5 mM final concentration of each salt was included in the cryoprotectant. Single crystals were mounted in a rayon-fibre loop and transferred to the cryoprotectant for a few seconds prior to flash freezing in liquid nitrogen.

2.2.3.8.6 Crystal screening

The crystals were mounted onto a single-axis goniometer and screened at the York Structural Biology Laboratory (YSBL) X-ray source to determine diffraction quality prior to data collection. Data collection was achieved using a 345 mm Mar Research image–plate detector on a Rigaku rotating anode RU-200 X-ray generator, with a Cu target operating at 50kV (100mA) and focusing X-ray optics (MSC). The quality of the crystal was judged by the diffraction pattern.

2.2.3.8.7 Data collection

Data sets were collected at the European Synchrotron Radiation Facility (ESRF) using beamline ID-14.2 and an ADSC Quantum-4 charge-coupled device detector. The cell index and collection strategy was determined using MOSFLM (Leslie, 1992). Owing to the long cell axis of HylP1 and HylP3, spot separation was achieved using an oscillation angle (θ) of 0.2° through an oscillation range (ψ) of 90°.

2.2.3.8.8 Data processing

All data were processed using MOSFLM (Leslie, 1992) and all other computing used the CCP4 suite (Collaborative Computational Project Number 4, 1994).

2.2.4 Assay methods for *Ct*CBM6

2.2.4.1 Affinity gel electrophoresis (AGE)

An AGE method using native polyacrylamide gels containing soluble ligand sugars was developed based on the method of Takeo (1984) and Tomme *et al.*, (1996).

A continuous gel system was used with the same gel apparatus (Appendix D8) as used for denaturing gels (Section 2.2.3.1). Gels contained 7.5 % acrylamide (Section 2.1.9.2.1) in 25 mM Tris/250 mM glycine buffer, pH 8.3, which was also the Running and Sample buffers (Section 2.1.9.2.3 and 2.1.9.2.4, respectively). For qualitative AGE, ligands were added at a 0.1 % (w/v) final concentration prior to polymerisation (Section 2.1.10). For quantitative AGE, a range of concentrations of oat spelt xylan (0.005-1.0 mg ml^{-1} final concentration; Section 2.1.10) were incorporated into the gel prior to polymerisation. Pure protein (5-10 μg) samples in 10 μl sample buffer, containing 5 % (v/v) glycerol and 0.0025 % bromophenol blue final concentration (Section 2.1.9.2.4), were electrophoresed at room temperature at 10 mA/gel for 3 h. Proteins were visualised by staining with 0.4 % (w/v) Coomassie Blue, 40 % (v/v) methanol, 10 % (v/v) glacial acetic acid for 10 min and the gel destained in 40 % (v/v) methanol, 10 % (v/v) glacial acetic acid (Sections 2.1.9.1.7 and 2.1.9.1.8, respectively). Gels with and without ligand were run in the same gel tank with identical samples loaded on each. BSA (5 μg) was used as a negative, non-interacting control (Section 2.1.9.2.2).

2.2.4.2 Isothermal titration calorimetry (ITC)

ITC measurements were carried out on a MicroCal Omega titration calorimeter. All solutions were thoroughly degassed by vacuum stirring prior to loading on the ITC machine. A 2 ml protein solution (0.1 – 0.2 mM) in 50 mM sodium phosphate buffer, pH 7.0, was carefully loaded to avoid bubbles, into the ITC machines sample cell using a 5 ml syringe connected to a 15 cm length of capillary tubing. Oligosaccharides (200 μl of 5 – 30 mM; section 2.1.10) or polysaccharides (200 μl of 2 – 4 mg ml^{-1}; section 2.1.10), dissolved in buffer from the final protein dialysis to minimise heats of dilution, was loaded into the ITC syringe, and the solution drawn up and down the syringe several times to ensure

no air bubbles were present, before placing it in the ITC machine. The machine was allowed to equilibrate before 25 successive aliquots of 10 µl of the titrant solution were injected at 3 min intervals into the sample cell from the syringe under continuous mixing at 400 rpm. Raw binding data were corrected for heats of dilution of both protein and ligands, obtained by independent titration experiments. Integrated heat effects, after correction for heats of dilution, were analysed by non-linear regression using a simple one site-binding model with the standard MicroCal ORIGIN software package. For each thermal titration, this yields estimates of the apparent number of binding sites on the protein (n), the K_a and the enthalpy of binding ($\Delta H°$). Other thermodynamic parameters ($\Delta G°$ and $T\Delta S°$) were calculated using the standard thermodynamic equation:

$$RT \ln K_a = \Delta G° = \Delta H° - T\Delta S°$$

Where, R = gas constant,

T = Temperature in degree absolute,

K_d = Equilibrium dissociation constant,

$\Delta G°$ = Change in free energy,

$\Delta H°$ = Enthalpy of binding,

$\Delta S°$ = Entropy of binding.

2.2.5 Assay methods for HylP1, HylP2 and HylP3

2.2.5.1 Basic biochemical characterisation of enzymes

2.2.5.1.1 General considerations for kinetic analysis

- All kinetics was performed using the same set of calibrated Gilson pipettes.
- The same aliquot of purified enzyme was used in kinetic experiments.
- All enzyme assays were performed in triplicate and the mean values were presented on the graphs with error bars to represent the range for each data set.

2.2.5.1.2 3,5-Dinitrosalicylic acid (DNSA) reducing sugar assay

The reducing sugar content of solutions was determined calorimetrically using DNSA (Miller, 1959).

Enzyme (HylP1 (0.17 μg), HylP2 (0.28 μg) or HylP3 (0.35 μg), diluted in 18.2 MΩ H_2O containing 1 mg ml^{-1} BSA (final concentration), was added to soluble dialysed sodium hyaluronate (4.0-0.5 mg ml^{-1}, final concentration; section 2.1.10) and an appropriate buffer (20 mM final concentration; Section 2.2.5.1.5), mixed by vortexing, and 150 μl aliquots removed at various time points (0-64 min) and added to 150 μl DNSA reagent (1 % (w/v) DNSA, 0.2 % (v/v) phenol, 1 % (w/v) NaOH, 0.002 % (w/v) glucose, 0.05 % (w/v) $NaSO_3$; section 2.1.9.6) to terminate the reaction. For 0 time, 150 μl of the reaction mix was removed prior to the addition of the enzyme. The tubes were then boiled for 20 min, placed on ice for 10 min and finally equilibrated to room temperature before reading the absorbencies. The absorbencies were determined at 570 nm in a 96-well plastic microtitre plate (200 μl of each reaction was measured) using a plate reader (Appendix D14). Reducing sugar release was determined from the standard curve and the amount of enzyme or substrate concentrations were adjusted to ensure that the linear rate of the reaction was determined. Standard curves (Figure 2.4) were produced each time an assay was carried out. The standard curves were generated by the addition of 0-1.0 mg ml^{-1} of glucose (final concentration; 10 mg ml^{-1} stock concentration) to 4.0-0.5 mg ml^{-1} (final concentration; section 2.1.10) of soluble sodium hyaluronate in 20 mM of an appropriate buffer (Section 2.2.5.1.5), containing 1 mg ml^{-1} BSA (final concentration) to a final volume of 150 μl. DNSA reagent (150 μl; Section 2.1.9.6) was added to the standard curve and the processed in the same way as the enzyme assays (post-incubation). Assays were carried out at 37 °C, unless otherwise stated.

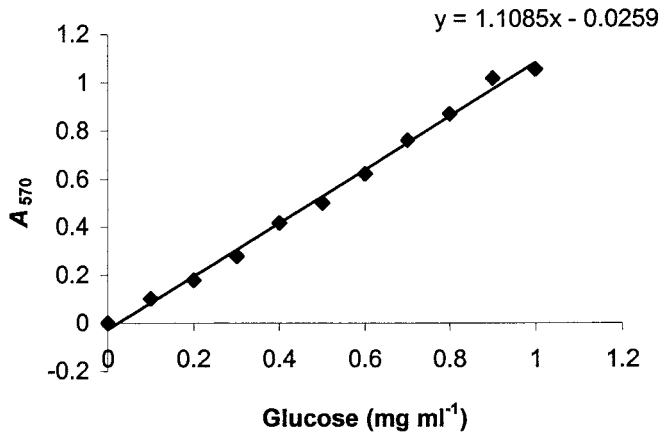


Figure 2.4: Typical standard curve for DNSA assays.
The curve was generated by the addition of 0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, + 1.0 mg ml^{-1} glucose (final concentration) to 4.0-0.5 mg/ml sodium hyaluronate (final concentration), 1 mg ml^{-1} BSA (final concentration) and 20 mM (final concentration) of an appropriate buffer, in a total volume of 150 µl. DNSA reagent (150 µl) was added, the tubes boiled for 20 min and placed on ice for 10 min. The tubes were finally equilibrated to room temperature before the A_{570} of the solution was determined.

2.2.5.1.3 Digests for PAGE analysis

Enzyme HylP1 (0.17 µg), HylP2 (0.28 µg) or HylP3 (0.35 µg), diluted in 18.2 MΩ H_2O, was added to 2 mg ml^{-1} dialysed sodium hyaluronate (final concentration; section 2.1.10) and 20 mM HEPES buffer, pH 7.0, mixed by vortexing and 150 µl aliquots removed at 0, 10, 20, 40, and 60 min. To terminate the reaction, each digest was boiled for 5 min before being analysed by PAGE (Section 2.2.5.2). For 0 time, 150 µl of the reaction mix was removed prior to the addition of the enzyme. At the same time points, an additional 150 µl of the reaction mix was removed and tested for reducing sugar content (Section 2.2.5.1.2) to ensure all enzymes were active.

2.2.5.1.4 Determination of K_M, k_{cat} and specific activity

For the determination of kinetic parameters of HylP1, HylP2 and HylP3 against sodium hyaluronate (2.1.10), varying substrate concentrations between 4.0-0.5 mg ml^{-1} (final concentrations) were chosen and assays performed as detailed in section 2.2.5.1.2. A standard curve was generated for each substrate concentration. To all assays, 0.17, 0.28 and 0.35 µg of HylP1, HylP2 and HylP3, respectively, and 20 mM Tris, pH 7.5, MES, pH 6.0 and sodium citrate, pH 5.5, respectively, were added. Aliquots of 150 µl were removed at 0, 0.5, 1, 2, 4, 8 and

16 min. For 0 time, 150 µl of the reaction mix was removed prior to the addition of the enzyme. All assays were carried out at 37 °C. In order to ascertain if these data reflected true Michaelis- Menten kinetics, a Lineweaver-Burk plot was plotted and if a straight line was observed the data were accepted as true.

2.2.5.1.5 Determination of pH optimum

For the determination of pH optimum of HylP1, HylP2 and HylP3, the substrate dialysed sodium hyaluronate (2 mg ml^{-1} final concentration; Section 2.1.10), enzyme HylP1 (0.17 µg), HylP2 (0.28 µg) and HylP3 (0.35 µg) and 20 mM of sodium citrate pH 3.0, 3.5, 4.0, 4.5, 5.0, 5.5, 5.8, 6.0, 6.2, 6.5, 7.0 were tested. BSA was not added to the reactions as precipitation occurred. Aliquots of 150 µl were removed at 0, 1, 2, 4, 8 and 16 min. For 0 time, 150 µl of the reaction mix was removed prior to the addition of the enzyme. Assays were performed as detailed in section 2.2.5.1.2.

2.2.5.1.6 Determination of temperature optimum and thermostability of HylP1, HylP2 and HylP3

Temperature optimum for HylP1, HylP2 and HylP3 were determined under standard assay conditions given in section 2.2.5.1.2, except the reaction was carried out at a range of temperatures (27, 37, 47, 57, 67 and 77 °C).

Thermostability of HylP1, HylP2 and HylP3 was assessed by pre-incubating the enzyme at 27, 37, 47, 57, 67, and 77 °C for 20 min prior to assaying the samples under standard conditions (Section 2.2.5.1.2).

HylP1 (0.17 µg), HylP2 (0.28 µg) and HylP3 (0.35 µg) were assayed against dialysed sodium hyaluronate (2 mg ml^{-1} final concentration; Section 2.1.10) and 20 mM Tris, pH 7.5, MES, pH 6.0 and sodium citrate, pH 5.5, respectively. Aliquots of 150 µl were removed at 0, 1, 2, 4, 8 and 16 min. For 0 time, 150 µl of the reaction mix was removed prior to the addition of the enzyme.

2.2.5.1.7 Determination of divalent ion requirement of HylP1, HylP2 and HylP3

The effects of divalent cations on activity was assessed using the standard reaction conditions (Section 2.2.5.1.2), except the substrate concentration tested was 2 mg ml^{-1} (final concentration; Section 2.1.10), 20 mM Tris, pH 7.5, MES, pH 6.0 and sodium citrate, pH 5.5, respectively, and divalent ions $CaCl_2$, $BaCl_2$, $CoCl_2$, $CuCl_2$, $MgCl_2$, $MnCl_2$, $NiCl_2$, $SrCl_2$, $ZnCl_2$ were all tested at 1 mM. Aliquots of 150 µl were removed at 0, 1, 2, 4, 8 and 16 min. For 0 time, 150 µl of the reaction mix was removed prior to the addition of the enzyme.

2.2.5.2 Mode of action of HylP1, HylP2 & HylP3 by PAGE analysis

The procedures employed for vertical-slab gel electrophoresis were essentially as described by Min & Cowman (1986) and Ikegami-Kawai & Takahasi (2002). Mini vertical slab gels (10.1 x 7.2 cm and 10.1 x 8.2 cm; appendix D8) were used with 10 sample wells (1.1 x 0.75 cm). The polyacrylamide gels contained 15.5 % acrylamide in 0.1 M Tris/Borate – 1 mM Na_2EDTA, pH 8.3 (TBE; Section 2.1.9.3.1). For the electrophoretic run, 10 µl of the hyaluronan digest samples (Section 2.2.5.1.3) were mixed with a one-fifth volume of 2M sucrose in TBE buffer (Section 2.1.9.3.5) and applied directly to the gel. Bromophenol Blue (BPB; 0.005 % in TBE buffer containing 0.3 M sucrose; Section 2.1.9.3.6) was used as a tracking dye and applied to a well with no sample. The reservoir buffer was TBE (10 x; Section 2.1.9.3.4) and was diluted 1:10 before use. In order to obtain a ladder-like series of bands, the gels were electrophoresed first at 250 V for 20 min, then at 500 V for 10 min, and additionally at 450 V for approximately 5 min or until the BPB tracking dye reached within 1 cm of the gel bottom. All the procedures of the electrophoretic run were carried out in a cold room.

A combined Alcian Blue and Silver staining protocol was employed for the detection of oligosaccharides. Immediately after electrophoresis, oligosaccharides were fixed in the gel matrix by soaking the gel in 0.05 % Alcian blue (Section 2.1.9.3.2 dissolved in 18.2 MΩ H_2O (~ 200 ml)) for 0.5 h in the dark. After destaining in 18.2 MΩ H_2O (~ 200 ml) for 0.5 h, the gel was subjected to Silver staining under normal room light, beginning at the oxidation step. Gels were soaked for 5 min in oxidiser solution (0.0034 M $K_2Cr_2O_7$, 0.0032 N HNO_3; ~ 200 ml; 2.1.9.3.3) while being shaken gently. Gels were then washed four times with 18.2 MΩ H_2O (~ 200 ml) for 30 sec until the yellow colour of the oxidiser

solution disappeared and soaked in Silver nitrate solution (0.012 M $AgNO_3$; ~ 200 ml; 2.1.9.3.3) for 30 min with gentle shaking. This was followed by two rapid rinses with image developer solution (0.28 M Na_2CO_3, 0.5 ml formaldehyde; ~ 200 ml; 2.1.9.3.3). A final portion (~ 200 ml) of the developer solution was added and shaken until a brown, smoky precipitate evolved of sufficient intensity. The development was stopped by discarding the developer solution and adding 1 % acetic acid (~ 100 ml; 2.1.9.3.3) for ~ 5 min. Gels were then washed twice with 18.2 MΩ H_2O for final storage.

2.2.5.3 Polysaccharide binding assays using SDS-PAGE

Qualitative binding of proteins HylP1, HylP2 and HylP3 to insoluble collagen was carried out as follows: pure protein (185, 309 and 390 µg, respectively) was mixed with 3 ml of 5 % (w/v) insoluble collagen in 50 mM HEPES buffer, pH 7.0, in a universal. The universals were incubated on ice for 1 h with orbital shaking at 50 rpm. The mixture was then filtered through a porosity grade 3 sintered glass filter (Pyrex) by vacuum suction. The insoluble collagen deposited on the filter was then washed three times by re-suspension in 3 ml of 50 mM HEPES buffer, pH 7.0, followed by vacuum suction. After the final wash, the insoluble collagen was scraped off the filter and re-suspended in 250 µl of 10 % (w/v) SDS. Any bound protein was then eluted from the insoluble collagen by boiling for 5 min. Samples (equivalent volumes of start material, all the wash fractions and eluant) were then analysed by SDS - PAGE (Section 2.2.3.1). Control experiments were carried out using BSA (50 µg) as a negative non-binding control.

Investigation of the interactions between *Clostridium thermocellum* Xyn11A CBM6 and its ligands.

3 Results

3.1 Introduction

The non-catalytic domains of modular xylanases that interact with polysaccharides are designated as carbohydrate-binding modules (CBMs) (Tomme *et al.*, 1995a; Boraston *et al.*, 1999; Coutinho and Henrissat, 1999). Although the majority of CBMs from plant cell wall (PCW) hydrolases characterised to date interact specifically with cellulose, modules have also been characterised that are able to bind to xylan (Tomme *et al.*, 1995a). Xylan, one of the major hemicelluloses in the PCW, consists of β-1,4-linked D-xylose units, which, unlike cellulose, are often substituted with arabinosyl, glucuronosyl, acetyl, uronyl and mannosyl residues (Hazlewood & Gilbert, 1993). CBMs that bind to xylan have so far only been found in enzymes involved in xylan degradation and are summarised in Table 3.1.

CBMs have been classified into 39 families to date, based on primary structure similarity (http://afmb.cnrs-mrs.fr/CAZY/; Coutinho and Henrissat, 1999). Ligand specificity can vary both between and within families, demonstrating considerable diversity in polysaccharide recognition in these protein modules (Coutinho and Henrissat, 1999).

Organism	CBM Family	Reference
Cellulomonas fimi	2b	Millward-Saddler *et al.*, 1994; Black *et al.*, 1995; Clarke *et al.*, 1996
Streptomyces thermoviolaceus		Kittur *et al.*, 2003
Rhodothermus marinus	4b	Nordberg *et al.*, 1997
Cellvibrio mixtus	6	Henshaw *et al.*, 2004; Pires *et al.*, 2004
Clostridium thermocellum		Fernandes *et al.*, 1999; Czjzek *et al.*, 2001
Clostridium stercorarium		Sun *et al.*, 1998; Boraston *et al.*, 2003
Streptomyces lividens	13	Dupont *et al.*, 1998
Streptomyces olivaceoviridis		Fujimoto *et al.*, 2002
*Cellvibrio japonicus**	15	Szabó *et al.*, 2001; Pell *et al.*, 2003
Clostridium thermocellum	22b	Charnock *et al.*, 2000; Xie *et al.*, 2001
Thermotoga maritima		Winterhalter *et al.*, 1995
*Cellvibrio japonicus**	35	Kellet *et al.*, 1990
Paenibacillus polymyxa	36	Gosalbes *et al.*, 1991

Table 3.1: Examples of characterised xylan-binding CBMs from different families. Based on data obtained from the CAZy classification system (**http://afmb.cnrs-mrs.fr/CAZY/**). * Denotes formerly *Pseudomonas cellulosa.*

Xyn11A (formerly XynU) is a modular enzyme from *C. thermocellum* strain YS. In addition to the two catalytic modules, a family 11 glycoside hydrolase (GH11) and a family 4 carbohydrate esterase (CE4), the xylanase contains a family 6 CBM (*Ct*CBM6; Fernandes *et al.*, 1999; Czjzek *et al.*, 2001; Figure 3.1).

Figure 3.1: Molecular architecture of *C. thermocellum* Xyn11A.
Xyn11A is a modular xylanase consisting of three discrete modules (shown as coloured boxes) joined by flexible linker regions (as shown by black lines). The N-terminal signal peptide is shown in black.

The anaerobic, thermophillic bacterium *C. thermocellum* synthesises a cluster of cellulases, xylanases, mannanases and other hydrolytic enzymes and CBMs that assemble into a large multi-enzyme complex termed the cellulosome (Béguin and Lemaire, 1996; Fernandes *et al.*, 1999; Shoham *et al.*, 1999; Carvalho *et al.*, 2003). Multienzyme PCW degrading complexes, such as the one found in *C. thermocellum*, are assembled by the interaction of two protein domains named

"cohesins" and "dockerins" (Carvalho *et al.*, 2003). Carvalho *et al.* (2003) have recently determined the 3D structure of the *C. thermocellum* dockerin-cohesin complex. The basis for the assembly of the cellulosome is the strong interaction between "cohesin" domains of the scaffold and the "dockerin" domains tethered to each enzyme (Béguin and Lemaire, 1996; Shoham *et al.*, 1999; Carvalho *et al.*, 2003). It is generally believed that assembly of the catalytic components into a complex enhances the synergistic interactions between enzymes with complementary activities, leading to more efficient PCW degradation (Fierobe *et al.*, 2002), whilst close association of this catalytic machine with the microbial cell wall enables the nutrients released to be preferentially utilised by the cellulosome-expressing host.

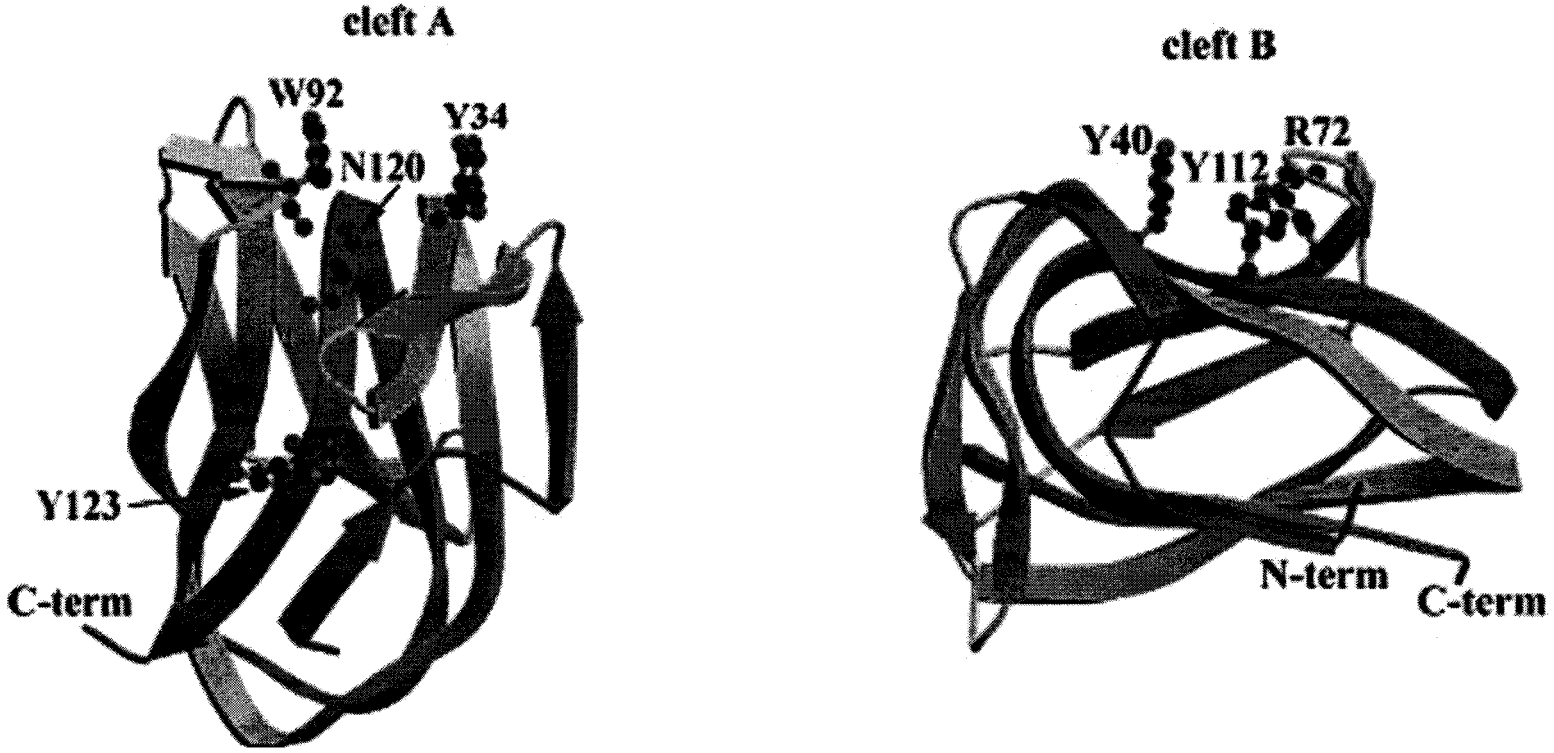


Figure 3.2: Ribbon representation of *Ct*CBM6 3D structure.
(A) View showing the ligand-binding cleft (cleft A) formed by the loops between the two β-sheets of the sandwich fold. The shallow binding cleft on top of the globular molecule is formed by Tyr-34 and Trp-92. (B) Ribbon representation of *Ct*CBM6 in a perpendicular view with respect to A, showing the second possible cleft (cleft B), which is obstructed by a short loop in *Ct*CBM6, situated on the concave face of the β-sheet sandwich. Taken from Czjzek *et al.*, 2001.

The structure of *Ct*CBM6 was determined by X-ray crystallography by Dr Czjzek (CNRS, Marseille). The protein forms a classic lectin-like β-jelly roll (Czjzek *et al.*, 2001; Figure 3.2) and displays a small shallow cleft, lined by two aromatic residues Tyr-34 and Trp-92 (Cleft A). However, when the fold was compared with CBM4 and CBM22, a second putative ligand-binding site becomes apparent on the concave face of the jelly roll (Figure 3.3; Cleft B). Cleft B strongly resembles the ligand-binding sites of CBM4 (Johnson *et al.*, 1996) and CBM22 (Charnock *et al.*, 2000), and two aromatic residues are located close to the surface, namely Tyr-40 and Tyr-112, the aromatic nature of which are relatively conserved in CBM6 (Tyr-40 is replaced by a tryptophan in a number of family 6 members; Figure 3.3), that could play a role in saccharide binding. Surprisingly, a proline residue of a neighbouring loop (residues 73-79) covers up this groove, making the surface aromatic residues apparently inaccessible to polymeric ligands.

This chapter will describe the interaction between CBM6 of *C. thermocellum* Xyn11A and its ligands. In order to study the role of individual conserved amino acids in the binding site, six mutants of *Ct*CBM6 were made by site-directed mutagenesis (SDM; Section 2.2.2.2). The properties of the mutants were compared with those of wild type protein using non-denaturing affinity gel electrophoresis (AGE; Section 2.2.4.1), isothermal titration calorimetry (ITC; Section 2.2.4.2) and NMR spectroscopy (performed by Dr M. Czjzek, CNRS Marseille). The data was used to determine the location of the ligand-binding site on the protein and led to the proposed model of ligand binding.

The results show that *Ct*CBM6 binds preferentially to xylan and xylooligosaccharides, interacts weakly with β-glucan and cellooligosaccharides and some soluble substituted forms of cellulose, but exhibits no detectable affinity for other mannose-containing hemicelluloses or pectins. ITC revealed that *Ct*CBM6 can accommodate up to five xylose residues, with binding driven mainly by enthalpic forces, partly offset by a negative entropy, in common with other soluble carbohydrate/protein interactions.

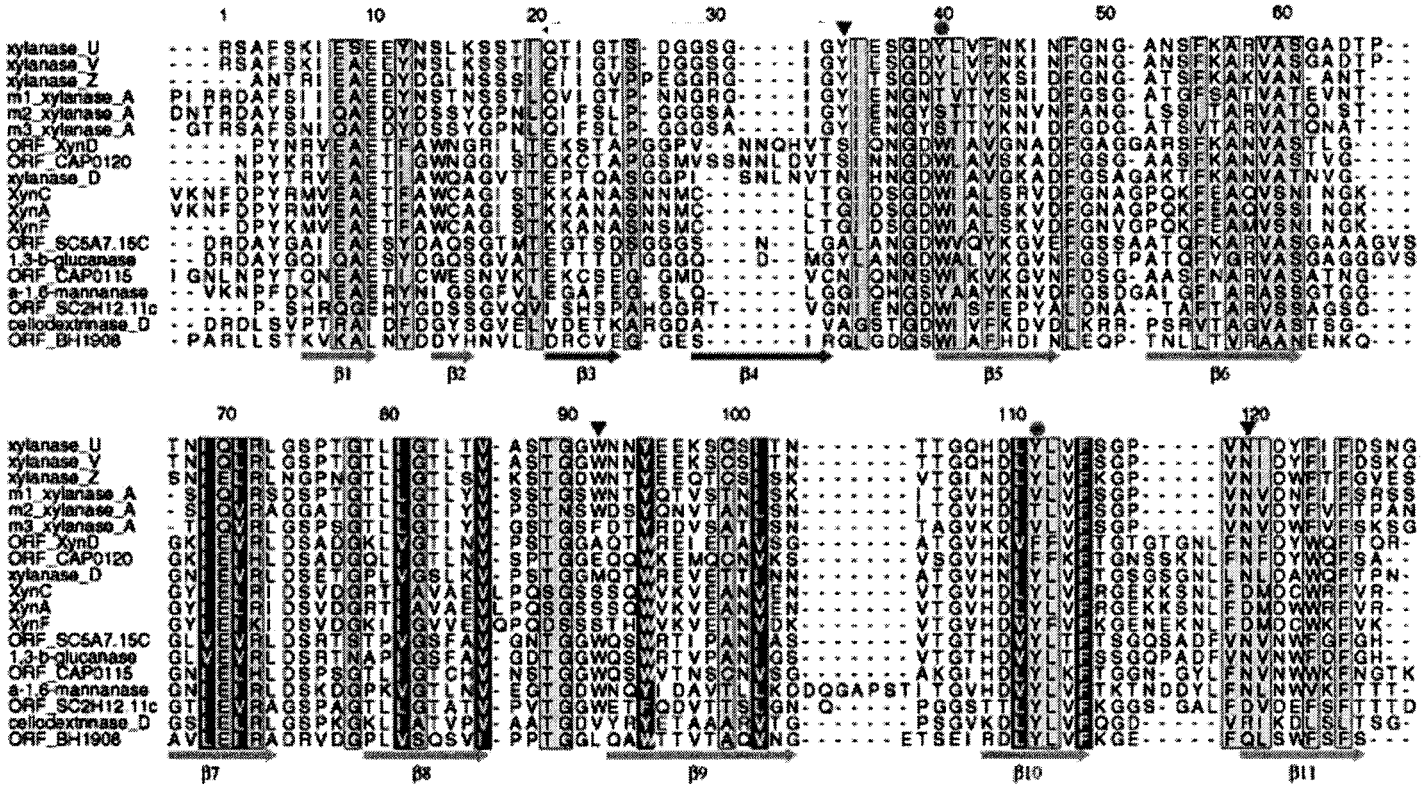


Figure 3.3: Sequence alignment of the carbohydrate-binding modules of family 6.
Sequence alignment of selected sequences from CBM6. The numbering corresponds to the sequence of CBM6 from *Clostridium thermocellum* Xyn11A (formerly XynU). The secondary structural elements (exclusively β-strands) are marked below the sequences in blue when part of the structural jelly roll fold and in green for two β-strands that are not part of the β-sandwich. The three residues, Tyr-34, Trp-92, and Asn-120, which have been shown to be involved in ligand binding, are marked by red arrowheads above the sequences. The pink circles above the sequences mark the position of two aromatic residues situated in cleft B. The alignment was performed by ClustalW. From top to bottom: xylanase U (Xyn11A from *C. thermocellum* (Swiss-Prot accession number O52780); xylanase V from *C. thermocellum* (Swiss-Prot accession number O52779); xylanase Z from *C. thermocellum* (Swiss-Prot accession number P10478); m-1, m-2, and m-3 of xylanase A from *C. stercorarium* strain NCIB11754 (GenBank™ accession number AF417638); ORF XynD from *Bacillus subtilis* 168 (Swiss-Prot accession number Q45071); ORF CAP0120 from *Clostridium acetobutylicum* (GenBank™ accession number AE001438); xylanase D from *Bacillus polymyxa* (Swiss-Prot accession number P45796); XynC from *Caldicellulosiruptor* sp. Rt69B.1 (Swiss-Prot accession number O52374); XynA from *Caldicellulosiruptor* sp. Tok7B.1 (Swiss-Prot accession number Q9X3P5); XynF from *Caldicellulosiruptor saccharolyticus* (Swiss-Prot accession number O30426); ORF SCA7.15C from *Streptomyces coelicolor* (Swiss-Prot accession number O88021); endo-1,3-β-glucanase from *Streptomyces sioyaensis* (GenBank™ accession number AF217415); ORF CAP0115 from *Clostridium acetobutylicum* (GenBank™ accession number AE001438); α-1,6-mannanase from *Bacillus circulans* TN31 (Swiss-Prot accession number Q9Z4P9); ORF SC2H12.11c from *Streptomyces coelicolor* (GenBank™ accession number AL359215); cellodextrinase D from *Microbispora bispora* (Swiss-Prot accession number Q59506); ORF BH1908 from *Bacillus halodurans* (GenBank™ accession number AP001513). Taken from Czjzek *et al.*, 2001.

3.2 Construction of *Ct*Xyn11A CBM6 and mutants

Plasmid p*Ct*Xyn11A, encoding *Ct*CBM6, was constructed previously (Fernandes *et al.*, 1999). Briefly, the region of the Xyn11A gene, encoding CBM6, was amplified by PCR, using the following primers that contain *Nde* I and *Xho* I restriction sites (engineered restriction sites are shown underlined): forward primer 5'CTCCATATgAAAATCgAATCTgAggAgTAC3' and reverse primer 3'CACCTCgAgTgTAggATTTACgCCATTCg5'. The amplified DNA was digested with *Nde* I / *Xho* I, and cloned into similarly restricted pET-21a vector to generate the plasmid p*Ct*Xyn11A, which encodes *Ct*CBM6 with a His_6 tag at its N-terminus.

To identify residues of *Ct*CBM6 involved in binding xylohexaose, NMR ligand titration experiments were carried out by Dr M. Czjzek, CNRS Marseille. The spectra of *Ct*CBM6 titrated with increasing amounts of xylohexaose showed that the protein and its ligand are in fast exchange between the free and bound states (Czjzek *et al.*, 2001). Thus, this allowed identification of the backbone *Ct*CBM6 amide groups that are affected by oligosaccharide binding. These residues include Tyr-34, Trp-92, and Asn-120, located in cleft A from the 3D structure (Figure 3.2). It is likely, therefore, that these residues are directly involved in ligand binding. To investigate the involvement of Tyr-34, Trp-92, and Asn-120 in ligand binding, these amino acids were each substituted with methionine (W92M) and alanine (Y34A and N120A). In addition, the residues Tyr-40, Arg-72, and Tyr-112, located in the second potential binding site cleft B (Figure 3.2), were replaced by methionine (Y40M, R72M and Y112M). Although NMR ligand titration experiments showed that none of the residues in cleft B were affected by the ligand, further investigation through SDM was required to determine whether cleft A was indeed the only ligand binding site in *Ct*CBM6. The primers used to generate these mutants, by the SDM method (Section 2.2.2.2), are given in Section 2.1.7.1. Mutants Y34A and N120A were made in-house and mutants R72M, Y40M, Y112M and W92M were made previously by Dr D.N. Bolam, Newcastle University.

PCR conditions were as given in section 2.2.2.1.1. The PCR products were transformed into *Escherichia coli* super competent XL10-Gold cells (Section 2.2.2.1.13.1), yielding ~50 colonies for each mutant. Plasmid DNA was purified from one colony (Section 2.2.2.1.6.2) and was sent for sequencing using the standard sequencing primers T7 promoter for Y34A and T7 terminator for N120A (Section 2.2.2.1.10). The data showed that the desired single codon mutations had been created successfully. Wild type and mutant forms of *Ct*CBM6 were transformed into *E. coli* BL21 chemically competent cells (DE3; section 2.2.2.1.13.1), and the recombinant *E. coli* strain was grown at 37 °C to mid-exponential phase. Expression of the CBMs was induced by the addition of IPTG at a final concentration of 1 mM, and the cultures incubated at 37 °C overnight (Section 2.2.1.2.3). Cells were harvested by centrifugation and periplasmic fractions were prepared as described in section 2.2.3.2.

The proteins were purified by immobilised metal affinity chromatography (IMAC) under non-denaturing conditions, using a Talon™/ chelating sepharose resin column (Section 2.2.3.4.2). An example of the purification is shown in Figure 3.4 (Section 2.2.3.1). Proteins were then dialysed against 50 mM sodium phosphate buffer, pH 7.0, at 12 °C and their concentrations were determined by Bradford's assay or A_{280} (Section 2.2.3.7.2 and 2.2.3.7.1, respectively) and subjected to AGE (Section 2.2.4.1) and ITC (Section 2.2.4.2).

3.3 Ligand binding studies

3.3.1 Qualitative affinity of *Ct*CBM6 and mutants to bind to soluble polysaccharides

Previous qualitative studies have shown that *Ct*CBM6 binds to insoluble and soluble polysaccharides (Fernandes *et al.*, 1999). However, the affinity of the protein for xylan, and the full range of other soluble ligands (Figure 3.5) that the CBM binds to, have not been defined. Semi-quantitative AGE was used to assess the capacity of *Ct*CBM6 to bind to a series of high molecular weight, soluble plant polysaccharides (Section 2.2.4.1; Takeo, 1984). In AGE, native protein samples are electrophoretically separated in polyacrylamide gels containing different concentrations of soluble polysaccharides. Reversible binding of the

protein to the ligand reduces its migration distance (mobility) due to the larger size of the protein/carbohydrate complex. The change in mobility is proportional to the polysaccharide concentration. The mobility of proteins that do not bind to the ligand in the gel would not be retarded even in the presence of high concentrations of the polysaccharide.

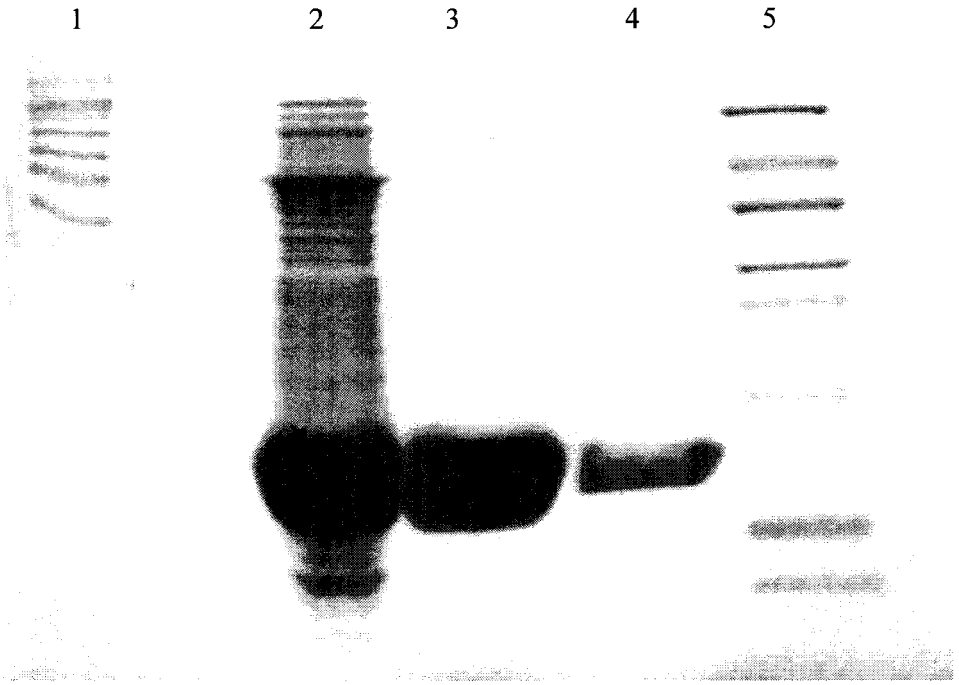


Figure 3.4: 15 % SDS-PAGE of *Ct*CBM6 hyper-expression and purification. CtCBM6 (wild type) was purified from E.coli cells harbouring pCtXyn11A by IMAC using Talon™ / chelating sepharose (Section 2.2.3.4.2). Lane 1: High molecular weight (HMW) markers (205, 116, 97, 84, 66, 55, 45, 36 kDa); lane 2: 20 µl periplasmic fraction (PPF) containing wild type CtCBM6; lane 3: 20 µl first elution (containing 100 mM final concentration of imidazole; Section 2.1.9.4.4); lane 4: 20 µl second elution; lane 5: Low molecular weight (LMW) markers (66, 45, 36, 29, 24, 20, 14.2, 6.5 kDa). Proteins were visualised by staining with Coomassie Blue and destained (Section 2.2.3.1.1). Expression levels of mutants Y34A, W92M, N120A, Y40M and R72M (data not shown) were similar to wild type CtCBM6, whereas the expression level of Y112M was considerably less.

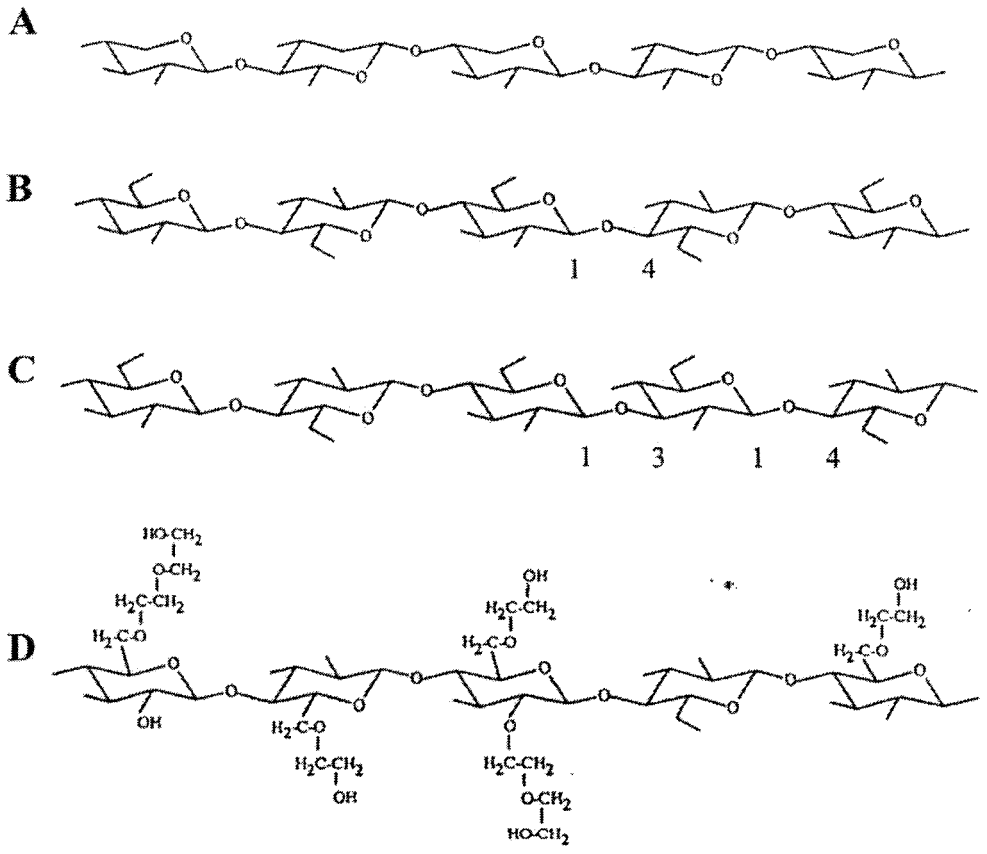


Figure 3.5: Structure of some of the polysaccharides used to investigate the binding specificity of CtCBM6 and mutants.
(A) β-1,4-xylooligosaccharide backbone of xylan; (B) β-1,4-linked glucan chain as found in cellulose; (C) β-1,3-1,4-glucan chain of barley β-glucan; and (D) β-1,4-linked glucan chain of hydroxyethylcellulose (HEC).

To compare the polysaccharide binding properties of wild type and mutants of *Ct*CBM6, the proteins were subjected to AGE using a range of soluble PCW structural polysaccharides (Section 2.1.10). In the use of AGE as a screen for binding specificity, aliquots of recombinant wild type *Ct*CBM6, the mutants (N120A, R72M, W92M, Y34A, Y40M and Y112M) and bovine serum albumin (BSA), chosen as a negative non-carbohydrate-binding control, were electrophoresed under native conditions at room temperature (Section 2.2.4.1) in gels containing 0.1 % (w/v) of the appropriate polysaccharide and a parallel gel containing no (potential) ligand.

Figures 3.6-3.9 show data from AGE experiments, in which the gels contained oat spelt xylan, birchwood xylan, rye arabinoxylan and wheat arabinoxylan, and the full data set is summarised in Table 3.2. The mutant N120A exhibited a greatly reduced affinity to oat spelt xylan, birchwood xylan, rye and wheat arabinoxylans, since the migration of N120A was not significantly retarded by the presence of these forms of xylan in the gel, while the native protein was strongly retarded by these polysaccharides (Figure 3.6-3.7). The mobility of W92M was affected more by the presence of varying forms of xylan, compared to N120A, but less than the wild type protein. Mutants Y34A, Y40M, R72M and Y112M appeared to show the same extent of retardation in the presence of these xylans as that of the wild type. This indicates that these mutations have not influenced the affinity of the protein for xylan under qualitative conditions. The wild type and some of the mutants showed a weak affinity for barley β-glucan (Y40M, R72M and Y112M), methyl cellulose (all mutants), hydroxyethyl cellulose (HEC; all mutants) and carboxymethylcellulose (CMC; Y112M). No retardation of *Ct*CBM6 and the mutants was observed with the gels containing carob and locust bean galactomannan, apple and lime pectin, sugar beet arabinan, pectin and potato galactan, CM-pachyman and rhamnogalacturonan. BSA was not retarded by the presence of any of the polysaccharides tested, indicating that the binding between *Ct*CBM6 and for example, oat spelt xylan, represented a true interaction of *Ct*CBM6 with the polysaccharide.

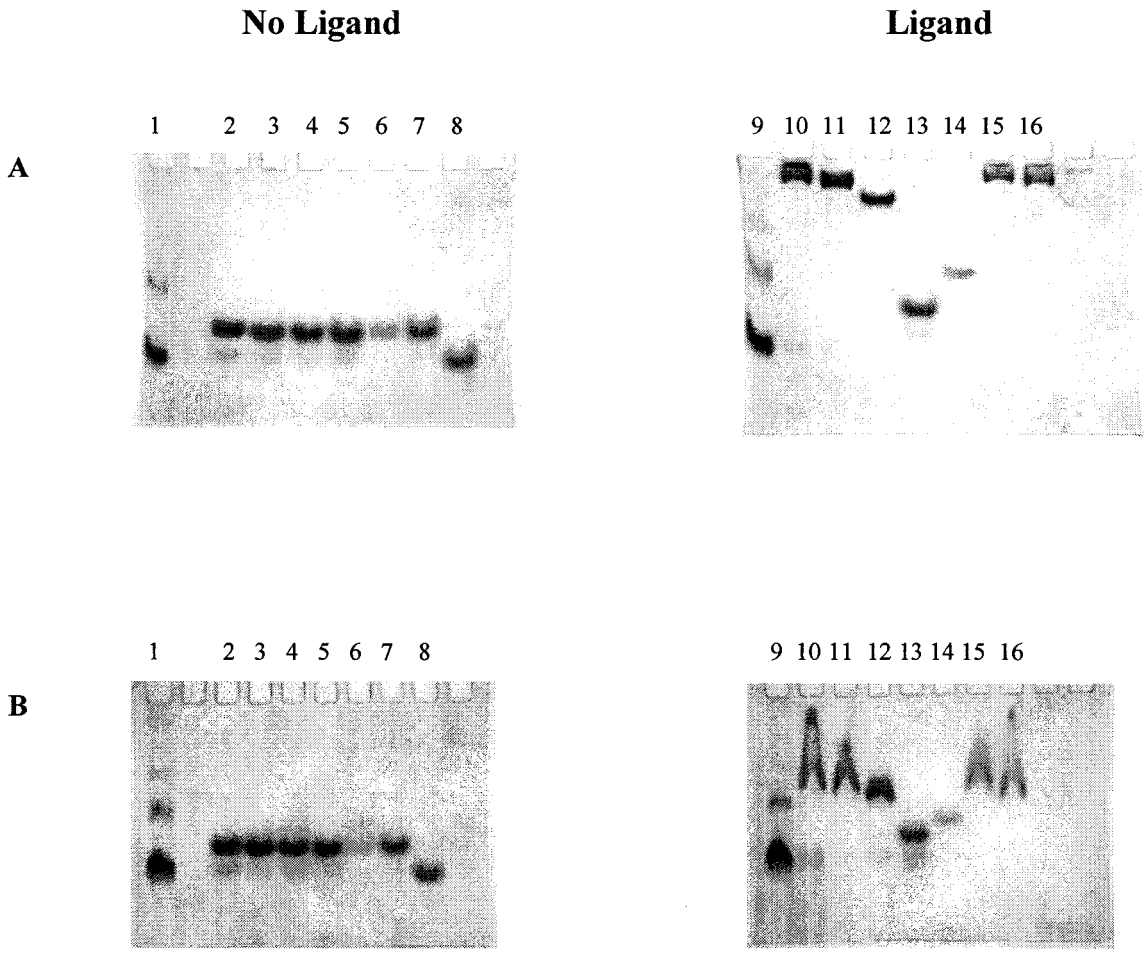


Figure 3.6: Qualitative AGE of wild type *Ct*CBM6 and the mutants in gels containing different ligands.
Wild type *Ct*CBM6, the mutants (Y34A, W92M, N120A, Y40M, R72M and Y112M) and BSA (non-binding control) were subjected to AGE, using a 7.5 % polyacrylamide gel in a Tris/glycine buffer system, pH 8.3, in the presence of different soluble polysaccharides at a final concentration of 0.1 % (w/v), as described in Section 2.2.4.1. The gels (with and without ligand) were electrophoresed for 3 h at 10 mA/gel. Proteins were visualised by staining with Coomassie Blue and destained (Section 2.2.3.1.1). (A) 0.1 % (w/v) Oat spelt xylan: No ligand: lane 1: 5 µg BSA; lane 2: 10 µl wild type *Ct*CBM6; lane 3: 10 µl Y112M; lane 4: 10 µl Y34A; lane 5: 10 µl N120A; lane 6: 10 µl W92M; lane 7: 10 µl Y40M; lane 8: 10 µl R72M. Ligand: lane 9: 5 µg BSA; lane 10: 10 µl wild type *Ct*CBM6; lane 11: 10 µl Y112M; lane 12: 10 µl Y34A; lane 13: 10 µl N120A; lane 14: 10 µl W92M; lane 15: 10 µl Y40M; lane 16: 10 µl R72M. (B) 0.1 % (w/v) Birchwood xylan: No ligand: Lane 1: 5 µg BSA; lane 2: 10 µl wild type *Ct*CBM6; lane 3: 10 µl Y112M; lane 4: 10 µl Y34A; lane 5: 10 µl N120A; lane 6: 10 µl W92M; lane 7: 10 µl Y40M; lane 8: 10 µl R72M. Ligand: lane 9: 5 µg BSA; lane 10: 10 µl wild type *Ct*CBM6; lane 11: 10 µl Y112M; lane 12: 10 µl Y34A; lane 13: 10 µl N120A; lane 14: 10 µl W92M; lane 15: 10 µl Y40M; lane 16: 10 µl R72M.

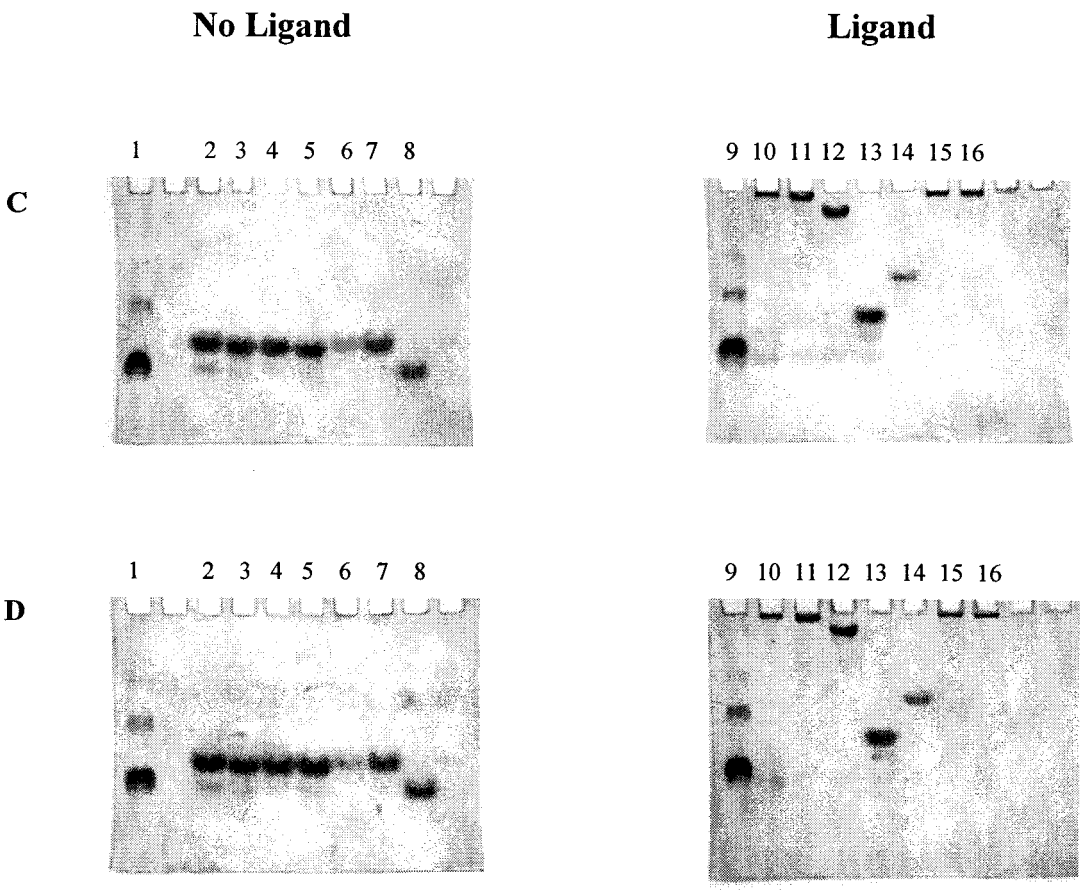


Figure 3.7: Qualitative AGE of wild type *Ct*CBM6 and the mutants in gels containing different ligands.
Wild type *Ct*CBM6, the mutants (Y34A, W92M, N120A, Y40M, R72M and Y112M) and BSA (non-binding control) were subjected to AGE, using a 7.5 % polyacrylamide gel in a Tris/glycine buffer system, pH 8.3, in the presence of different soluble polysaccharides at a final concentration of 0.1 % (w/v), as described in Section 2.2.4.1. The gels (with and without ligand) were electrophoresed for 3 h at 10 mA/gel. Proteins were visualised by staining with Coomassie Blue and destained (Section 2.2.3.1.1). (C) 0.1 % (w/v) Wheat arabinoxylan: No ligand: lane 1: 5 µg BSA; lane 2: 10 µl wild type *Ct*CBM6; lane 3: 10 µl Y112M; lane 4: 10 µl Y34A; lane 5: 10 µl N120A; lane 6: 10 µl W92M; lane 7: 10 µl Y40M; lane 8: 10 µl R72M. Ligand: lane 9: 5 µg BSA; lane 10: 10 µl wild type *Ct*CBM6; lane 11: 10 µl Y112M; lane 12: 10 µl Y34A; lane 13: 10 µl N120A; lane 14: 10 µl W92M; lane 15: 10 µl Y40M; lane 16: 10 µl R72M. (D) 0.1 % (w/v) Rye arabinoxylan: No ligand: Lane 1: 5 µg BSA; lane 2: 10 µl wild type *Ct*CBM6; lane 3: 10 µl Y112M; lane 4: 10 µl Y34A; lane 5: 10 µl N120A; lane 6: 10 µl W92M; lane 7: 10 µl Y40M; lane 8: 10 µl R72M. Ligand: lane 9: 5 µg BSA; lane 10: 10 µl wild type *Ct*CBM6; lane 11: 10 µl Y112M; lane 12: 10 µl Y34A; lane 13: 10 µl N120A; lane 14: 10 µl W92M; lane 15: 10 µl Y40M; lane 16: 10 µl R72M.

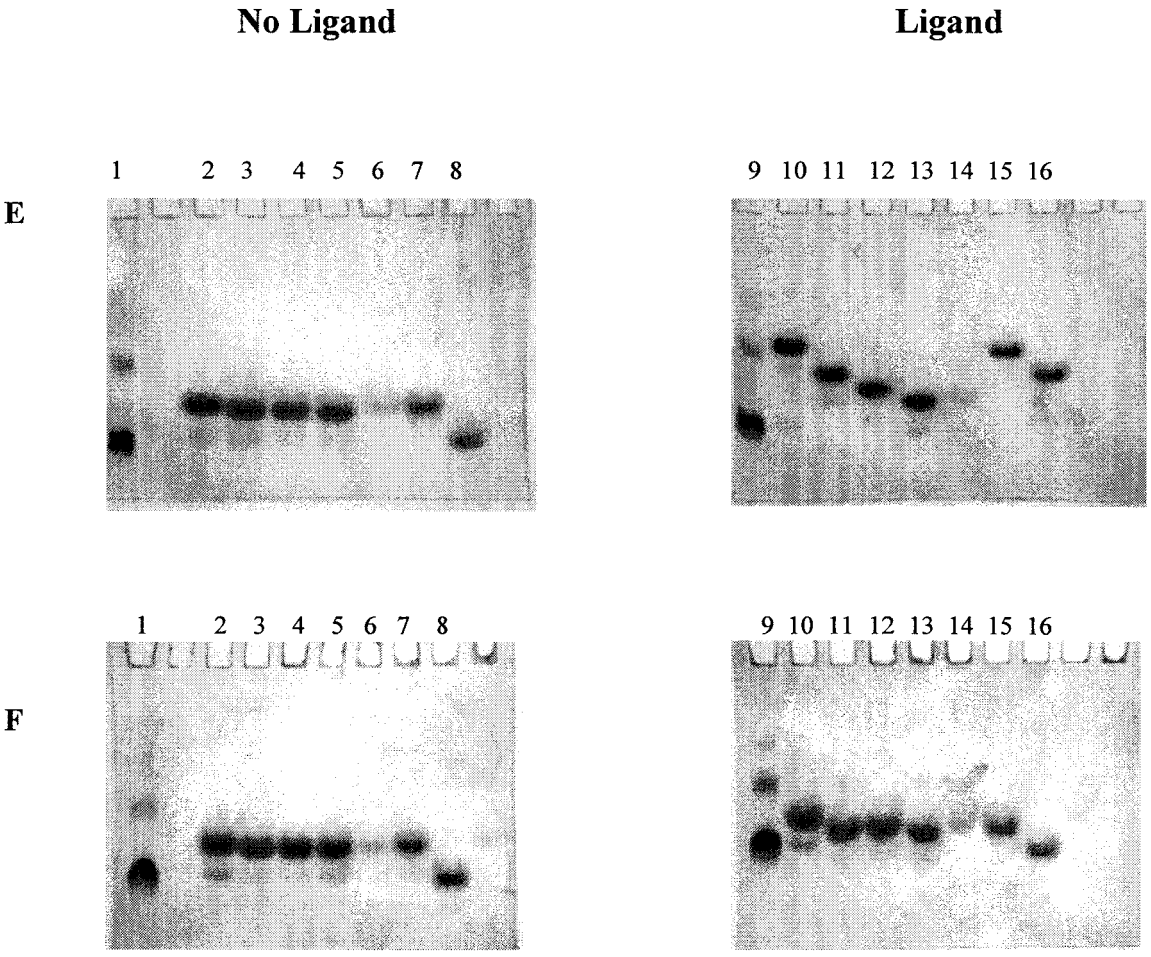


Figure 3.8: Qualitative AGE of wild type *Ct*CBM6 and the mutants in gels containing different ligands.
Wild type *Ct*CBM6, the mutants (Y34A, W92M, N120A, Y40M, R72M and Y112M) and BSA (non-binding control) were subjected to AGE, using a 7.5 % polyacrylamide gel in a Tris/glycine buffer system, pH 8.3, in the presence of different soluble polysaccharides at a final concentration of 0.1 % (w/v), as described in Section 2.2.4.1. The gels (with and without ligand) were electrophoresed for 3 h at 10 mA/gel. Proteins were visualised by staining with Coomassie Blue and destained (Section 2.2.3.1.1). (E) 0.1 % (w/v) Barley β-glucan: No ligand: lane 1: 5 μg BSA; lane 2: 10 μl wild type *Ct*CBM6; lane 3: 10 μl Y112M; lane 4: 10 μl Y34A; lane 5: 10 μl N120A; lane 6: 10 μl W92M; lane 7: 10 μl Y40M; lane 8: 10 μl R72M. Ligand: lane 9: 5 μg BSA; lane 10: 10 μl wild type *Ct*CBM6; lane 11: 10 μl Y112M; lane 12: 10 μl Y34A; lane 13: 10 μl N120A; lane 14: 10 μl W92M; lane 15: 10 μl Y40M; lane 16: 10 μl R72M. (F) 0.1 % (w/v) Carboxymethylcellulose (CMC): No ligand: Lane 1: 5 μg BSA; lane 2: 10 μl wild type *Ct*CBM6; lane 3: 10 μl Y112M; lane 4: 10 μl Y34A; lane 5: 10 μl N120A; lane 6: 10 μl W92M; lane 7: 10 μl Y40M; lane 8: 10 μl R72M. Ligand: lane 9: 5 μg BSA; lane 10: 10 μl wild type *Ct*CBM6; lane 11: 10 μl Y112M; lane 12: 10 μl Y34A; lane 13: 10 μl N120A; lane 14: 10 μl W92M; lane 15: 10 μl Y40M; lane 16: 10 μl R72M.

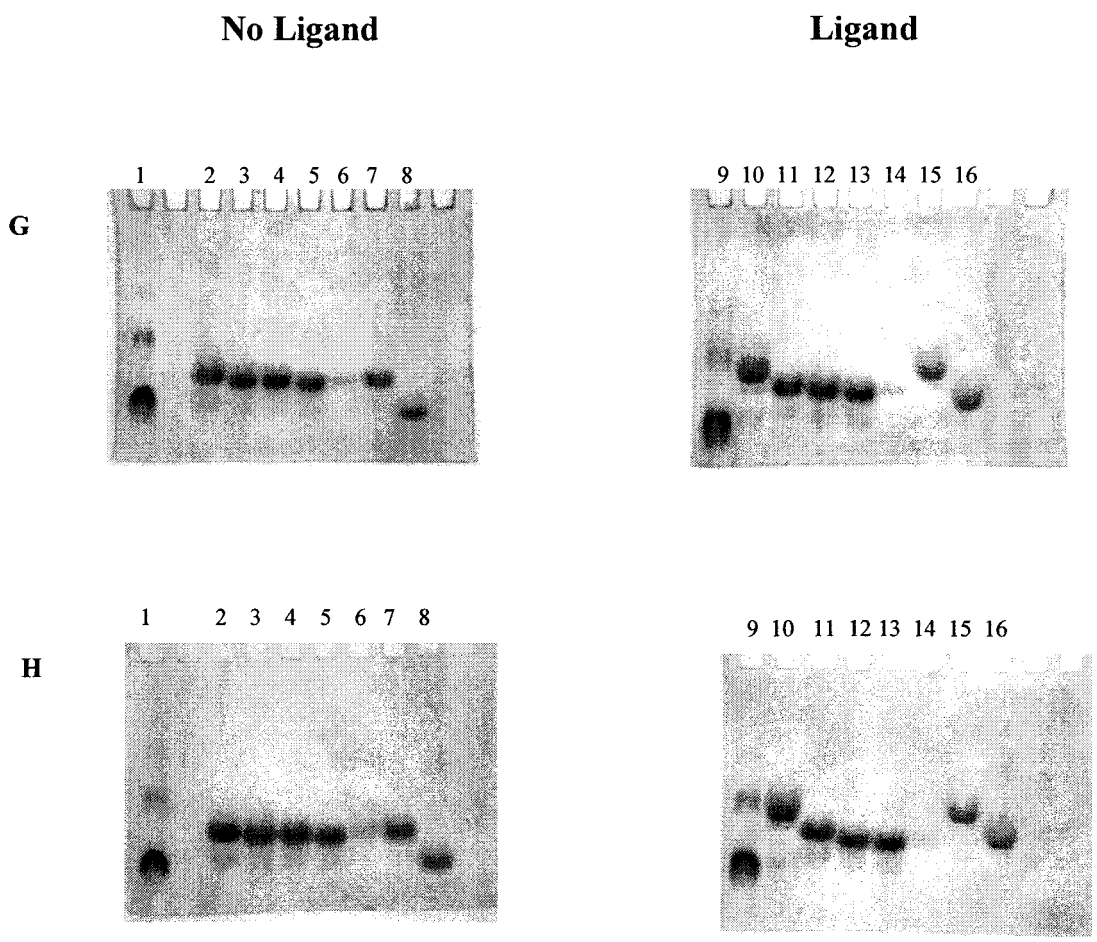


Figure 3.9: Qualitative AGE of wild type *Ct*CBM6 and the mutants in gels containing different ligands.
Wild type *Ct*CBM6, the mutants (Y34A, W92M, N120A, Y40M, R72M and Y112M) and BSA (non-binding control) were subjected to AGE, using a 7.5 % polyacrylamide gel in a Tris/glycine buffer system, pH 8.3, in the presence of different soluble polysaccharides at a final concentration of 0.1 % (w/v), as described in Section 2.2.4.1. The gels (with and without ligand) were electrophoresed for 3 h at 10 mA/gel. Proteins were visualised by staining with Coomassie Blue and destained (Section 2.2.3.1.1). (G) 0.1 % (w/v) Hydroxyethyl cellulose (HEC): No ligand: lane 1: 5 µg BSA; lane 2: 10 µl wild type *Ct*CBM6; lane 3: 10 µl Y112M; lane 4: 10 µl Y34A; lane 5: 10 µl N120A; lane 6: 10 µl W92M; lane 7: 10 µl Y40M; lane 8: 10 µl R72M. Ligand: lane 9: 5 µg BSA; lane 10: 10 µl wild type *Ct*CBM6; lane 11: 10 µl Y112M; lane 12: 10 µl Y34A; lane 13: 10 µl N120A; lane 14: 10 µl W92M; lane 15: 10 µl Y40M; lane 16: 10 µl R72M. (H) 0.1 % (w/v) Methyl cellulose: No ligand: Lane 1: 5 µg BSA; lane 2: 10 µl wild type *Ct*CBM6; lane 3: 10 µl Y112M; lane 4: 10 µl Y34A; lane 5: 10 µl N120A; lane 6: 10 µl W92M; lane 7: 10 µl Y40M; lane 8: 10 µl R72M. Ligand: lane 9: 5 µg BSA; lane 10: 10 µl wild type *Ct*CBM6; lane 11: 10 µl Y112M; lane 12: 10 µl Y34A; lane 13: 10 µl N120A; lane 14: 10 µl W92M; lane 15: 10 µl Y40M; lane 16: 10 µl R72M.

Soluble Polysaccharides	Major linkage of polysaccharide backbone	WT	Y34A	N120A	W92M	Y40M	R72M	Y112M
Oat Spelt Xylan	β-1,4-xylopyranose	+++	++	+	++	+++	+++	+++
Birchwood Xylan	β-1,4-xylopyranose	++	++	+	+	++	++	++
Rye Arabinoxylan	β-1,4-xylopyranose	+++	++	+	++	+++	+++	+++
Wheat Arabinoxylan	β-1,4-xylopyranose	+++	++	+	++	+++	+++	+++
Barley β Glucan	β-1,4-glucopyranose; some β-1,3-glucopyranose	+	-	-	-	+	+	+
Locust Bean Galactomannan	β-1,4-mannopyranose	-	-	-	-	-	-	-
Carob Galactomannan	β-1,4-mannopyranose	-	-	-	-	-	-	-
Hydroxyethyl Cellulose (HEC)	β-1,4-glucopyranose	+	+	+	+	+	+	+
Methyl Cellulose	β-1,4-glucopyranose	+	+	+	+	+	+	+
Carboxymethylcellulose (CMC)	β-1,4-glucopyranose	+	-	-	-	-	-	+
Lime Pectin	Mixture of polysaccharides	-	-	-	-	-	-	-
Apple Pectin	Mixture of polysaccharides	-	-	-	-	-	-	-
Sugar Beet Arabinan	α-1,5-arabinofuranose	-	-	-	-	-	-	-
Potato Galactan	β-1,4-galactopyranose	-	-	-	-	-	-	-
Pectin Galactan	β-1,4-galactopyranose	-	-	-	-	-	-	-
CM-Pachyman	1,3-β-linked D-glucosyl residues	-	-	-	-	-	-	-
Rhamnogalacturonan	α-1,4- galactopyranose	-	-	-	-	-	-	-

Table 3.2: Semi-quantitative affinities of *Ct*CBM6 (wild type) and mutants for various soluble plant cell wall polysaccharides.
*Ct*CBM6 (WT) and mutants (Y34A, N120A, W92M, Y40M, R72M and Y112M) were subjected to AGE (Section 2.2.4.1) in polyacrylamide gels containing different plant cell wall polysaccharides (0.1 % (w/v) final concentration). Maximum, significant, slight, and no retardation of the protein by the polsaccharide is indicated by +++, ++, +, and -, respectively.

To summarise, these results indicate that *Ct*CBM6 is primarily a CBM that binds to xylan of varying degrees of substitution.

3.3.2 Quantitative affinity of *Ct*CBM6 and mutants for soluble polysaccharides

To quantify the affinity of *Ct*CBM6 and mutants for oat spelt xylan, quantitative AGE was used. *Ct*CBM6 and mutants were subjected to gels with and without various different concentrations of oat spelt xylan. Each pair of gels (presence and absence of ligand, respectively) were run at 10 mA for the same duration (3 h; section 2.2.4.1). The distance moved by the wild type and mutants of *Ct*CBM6 reduced as the concentration [C] of xylan increased from 0 – 1.0 mg/ml, whereas the migration of the control BSA was not affected (Figures 3.10-3.13). The migration distance of *Ct*CBM6 and mutants in the presence and absence of ligand was then measured after the electrophoresed gels had been stained and destained (Section 2.2.3.1.1). Following destaining of each gel, it was noted that a second smaller faint band was also visible. When analysing the degree of retardation of CBM6 (WT) and mutants, the prominent higher MW band for each protein was used. CBMs, in general, are notorious for not running according to their MW, which may explain the appearance of a second band, or it may be possible that there is a proportion of the protein solution that is unable to interact with the ligand present in the gel and is therefore not retarded.

In AGE, the decrease in mobility of proteins in the presence of soluble ligand is proportional to the concentration of the ligand added. Thus, the dissociation and association constants of CBMs for soluble polysaccharides can be calculated by plotting the change in mobility as a function of ligand concentrations (Takeo, 1984). The relative mobility, which is used to calculate the dissociation constant, is expressed as the migration distance of the protein sample versus the migration distance of a non-binding control. Migration distances were measured, in cm, relative to the top of the gel (base of the wells). The relative migration distance is the ratio of the samples' migration distance (d) to that of the reference substance (D), BSA (d/D). In this way, the relative mobility of wild type and mutants of *Ct*CBM6 in the presence (r) and absence (R) of oat spelt xylan were obtained.

In the case of a protein-ligand complex where the protein migrates into the polyacrylamide gel, even at high concentrations of ligand, the general affinity equation (Equation 1; Takeo, 1984) was used to obtain a value for the dissociation constant (K_d).

$$1 / (R - r) = 1 / (R - R_c) \,.\, (1 + K_d / [C]) \qquad \text{Equation 1}$$

where, R is the relative migration distance (d/D) of the free protein (*Ct*CBM6), i.e. that obtained in the absence of the macromolecular ligand,
r is the migration distance (d/D) of the protein (*Ct*CBM6 and mutants) in the presence of ligands,
[C] is the total concentration of ligand (%),
K_d is the dissociation constant of protein for the ligand given as a % concentration ($C^{\%}$),
R_c is the relative migration distance of the complex between protein and affinity ligand, i.e. that value obtained in the presence of an excess amount of affinity ligand with all protein molecules binding the ligand.

When the general affinity equation (Equation 1) is plotted taking [C] % as the x axis and 1 /(R- r) as the y axis, a straight line is obtained. The equation of the line is obtained from the reciprocal graph and taking $y = 0$, the negative value of the dissociation constant (K_d) is equal to x and can be calculated. The association constant (K_a) for the interaction was obtained by taking the reciprocal of K_d ($1/K_d$).

The general affinity equation (Equation 1) was used to determine the dissociation and association constants for the interaction between wild type *Ct*CBM6 and the mutants with oat spelt xylan. Data, presented in Table 3.3, show that in the presence of a high excess of oat spelt xylan, wild type *Ct*CBM6 and the mutants still had some mobility (Figure 3.13), probably due to the incomplete molecular sieving effect of the polyacrylamide gel (Takeo, 1984). Figure 3.14 displays a reciprocal plot of relative mobility against ligand concentration for the wild type

CBM and the K_d values, estimated from the plot of 1/(R-r) against [C], are given in Table 3.3.

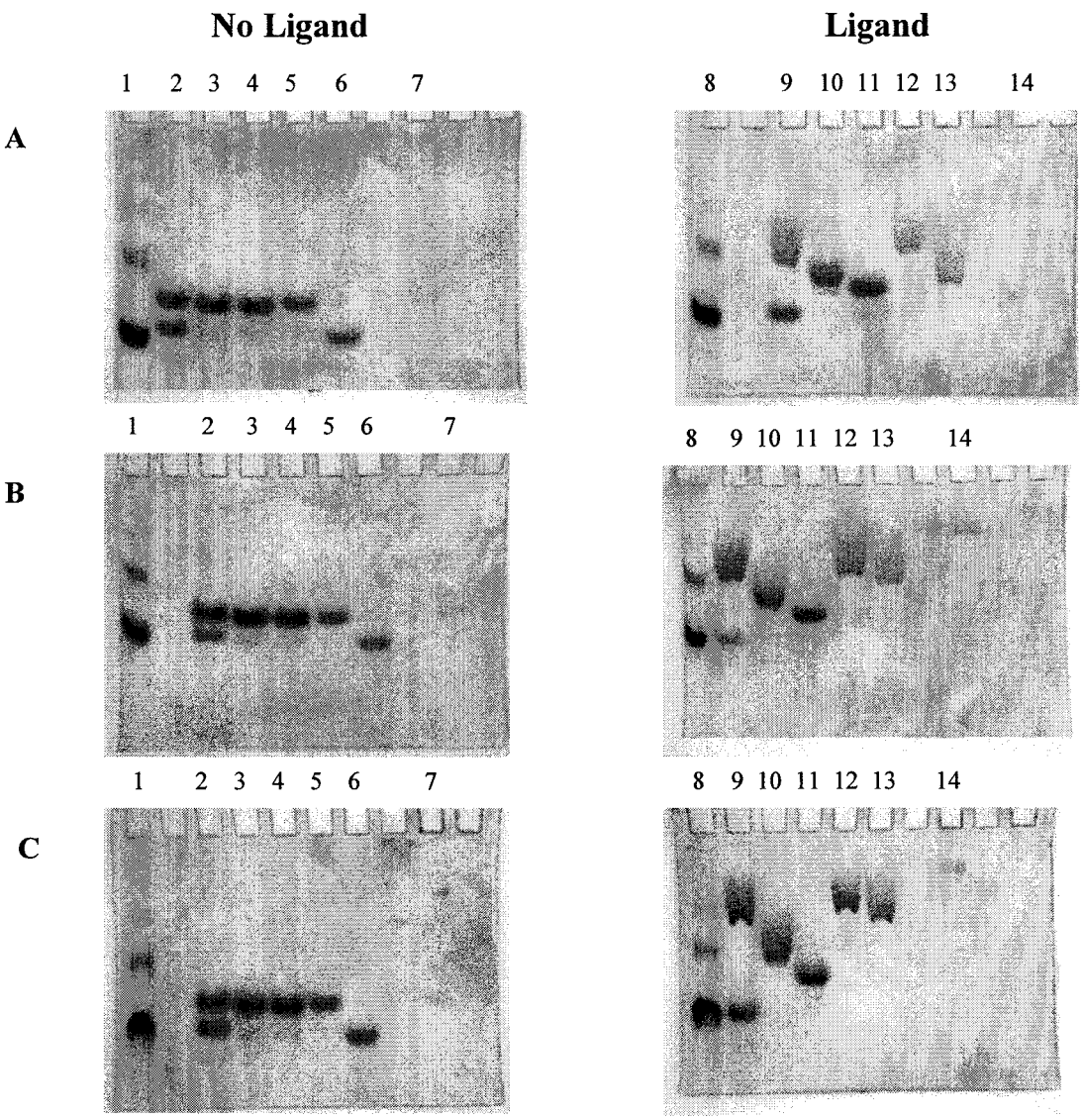


Figure 3.10: Quantitative AGE of wild type *Ct*CBM6 and the mutants in gels containing different concentrations of oat spelt xylan.
Wild type *Ct*CBM6, the mutants (Y34A, W92M, N120A, Y40M, R72M and Y112M) and BSA (non-binding control) were subjected to AGE, using a 7.5 % polyacrylamide gel in a Tris/glycine buffer system, pH 8.3, in the presence of different concentrations (final concentrations) of soluble oat spelt xylan, as described in Section 2.2.4.1. The gels (with and without ligand) were electrophoresed for 3 h at 10 mA/gel. Proteins were visualised by staining with Coomassie Blue and destained (Section 2.2.3.1.1). (A) 0.005 mg/ml: No ligand: lane 1: 5 µg BSA; lane 2: 10 µl wild type *Ct*CBM6; lane 3: 10 µl Y95F; lane 4: 10 µl Y34A; lane 5: 10 µl Y40M; lane 6: 10 µl R72M; lane 7: 10 µl Y112M. Ligand: lane 8: 5 µg BSA; lane 9: 10 µl wild type *Ct*CBM6; lane 10: 10 µl Y95F; lane 11: 10 µl Y34A; lane 12: 10 µl Y40M; lane 13: 10 µl R72M; lane 14: 10 µl Y112M. (B) 0.01 mg/ml: No ligand: Lane 1: 5 µg BSA; lane 2: 10 µl wild type *Ct*CBM6; lane 3: 10 µl Y95F; lane 4: 10 µl Y34A; lane 5: 10 µl Y40M; lane 6: 10 µl R72M; lane 7: 10 µl Y112M. Ligand: lane 8: 5 µg BSA; lane 9: 10 µl wild type *Ct*CBM6; lane 10: 10 µl Y95F; lane 11: 10 µl Y34A; lane 12: 10 µl Y40M; lane 13: 10 µl R72M; lane 14: 10 µl Y112M. (C) 0.015 mg/ml: No ligand: Lane 1: 5 µg BSA; lane 2: 10 µl wild type *Ct*CBM6; lane 3: 10 µl Y95F; lane 4: 10 µl Y34A; lane 5: 10 µl Y40M; lane 6: 10 µl R72M; lane 7: 10 µl Y112M. Ligand: lane 8: 5 µg BSA; lane 9: 10 µl wild type *Ct*CBM6; lane 10: 10 µl Y95F; lane 11: 10 µl Y34A; lane 12: 10 µl Y40M; lane 13: 10 µl R72M; lane 14: 10 µl Y112M.

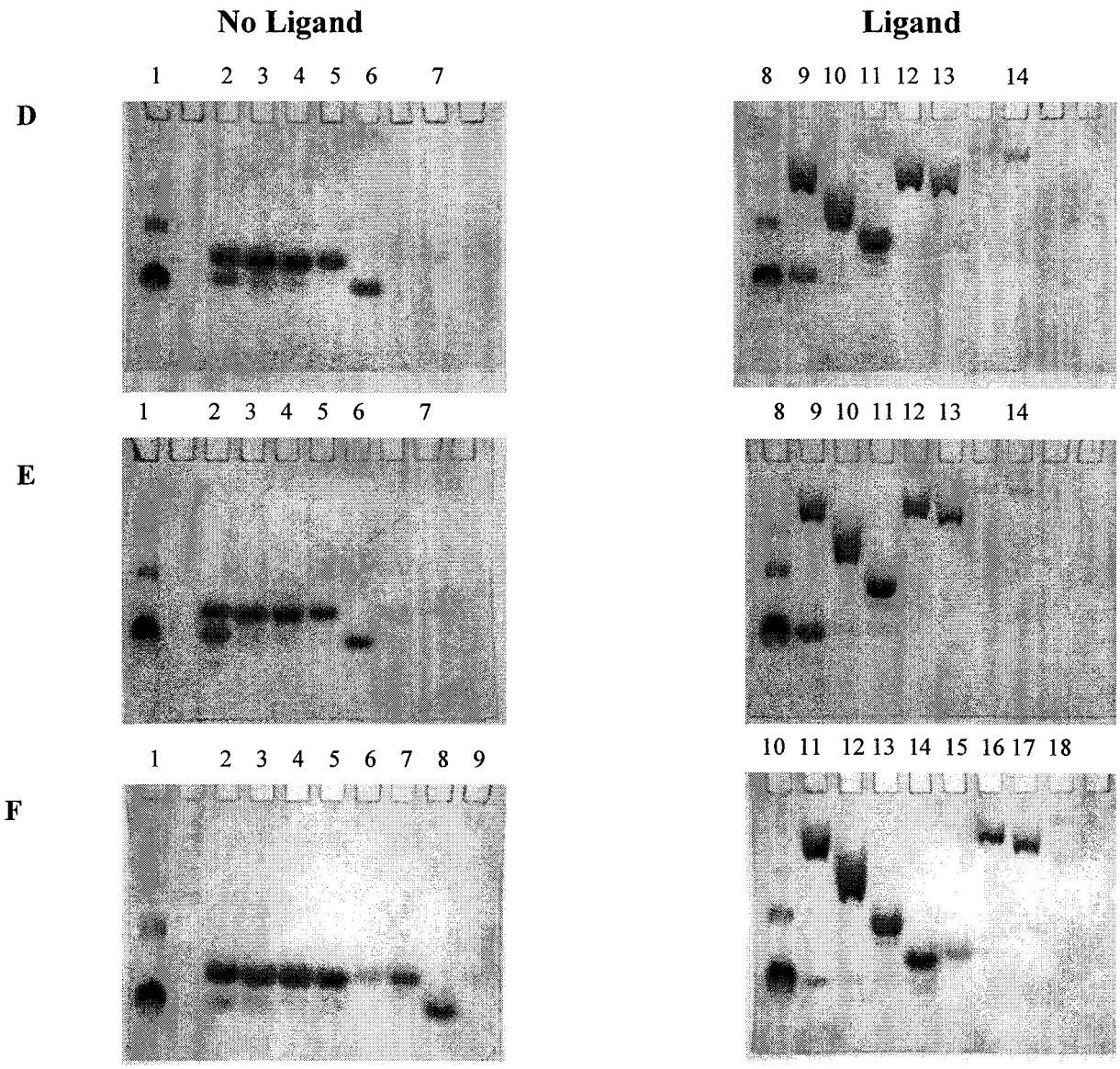


Figure 3.11: Quantitative AGE of wild type *Ct*CBM6 and the mutants in gels containing different concentrations of oat spelt xylan.

Wild type *Ct*CBM6, the mutants (Y34A, W92M, N120A, Y40M, R72M and Y112M) and BSA (non-binding control) were subjected to AGE, using a 7.5 % polyacrylamide gel in a Tris/glycine buffer system, pH 8.3, in the presence of different concentrations (final concentrations) of soluble oat spelt xylan, as described in Section 2.2.4.1. The gels (with and without ligand) were electrophoresed for 3 h at 10 mA/gel. Proteins were visualised by staining with Coomassie Blue and destained (Section 2.2.3.1.1). (D) 0.02 mg/ml: No ligand: lane 1: 5 µg BSA; lane 2: 10 µl wild type *Ct*CBM6; lane 3: 10 µl Y95F; lane 4: 10 µl Y34A; lane 5: 10 µl Y40M; lane 6: 10 µl R72M; lane 7: 10 µl Y112M. Ligand: lane 8: 5 µg BSA; lane 9: 10 µl wild type *Ct*CBM6; lane 10: 10 µl Y95F; lane 11: 10 µl Y34A; lane 12: 10 µl Y40M; lane 13: 10 µl R72M; lane 14: 10 µl Y112M. (E) 0.03 mg/ml: No ligand: Lane 1: 5 µg BSA; lane 2: 10 µl wild type *Ct*CBM6; lane 3: 10 µl Y95F; lane 4: 10 µl Y34A; lane 5: 10 µl Y40M; lane 6: 10 µl R72M; lane 7: 10 µl Y112M. Ligand: lane 8: 5 µg BSA; lane 9: 10 µl wild type *Ct*CBM6; lane 10: 10 µl Y95F; lane 11: 10 µl Y34A; lane 12: 10 µl Y40M; lane 13: 10 µl R72M; lane 14: 10 µl Y112M. (F) 0.04 mg/ml: No ligand: Lane 1: 5 µg BSA; lane 2: 10 µl wild type *Ct*CBM6; lane 3: 10 µl Y95F; lane 4: 10 µl Y34A; lane 5: 10 µl N120A; lane 6: 10 µl W92M; lane 7: 10 µl Y40M; lane 8: 10 µl R72M; lane 9: 10 µl Y112M. Ligand: lane 10: 5 µg BSA; lane 11: 10 µl wild type *Ct*CBM6; lane 12: 10 µl Y95F; lane 13: 10 µl Y34A; lane 14: 10 µl N120A; lane 15: 10 µl W92M; lane 16: 10 µl Y40M; lane 17: 10 µl R72M; lane 18: 10 µl Y112M.

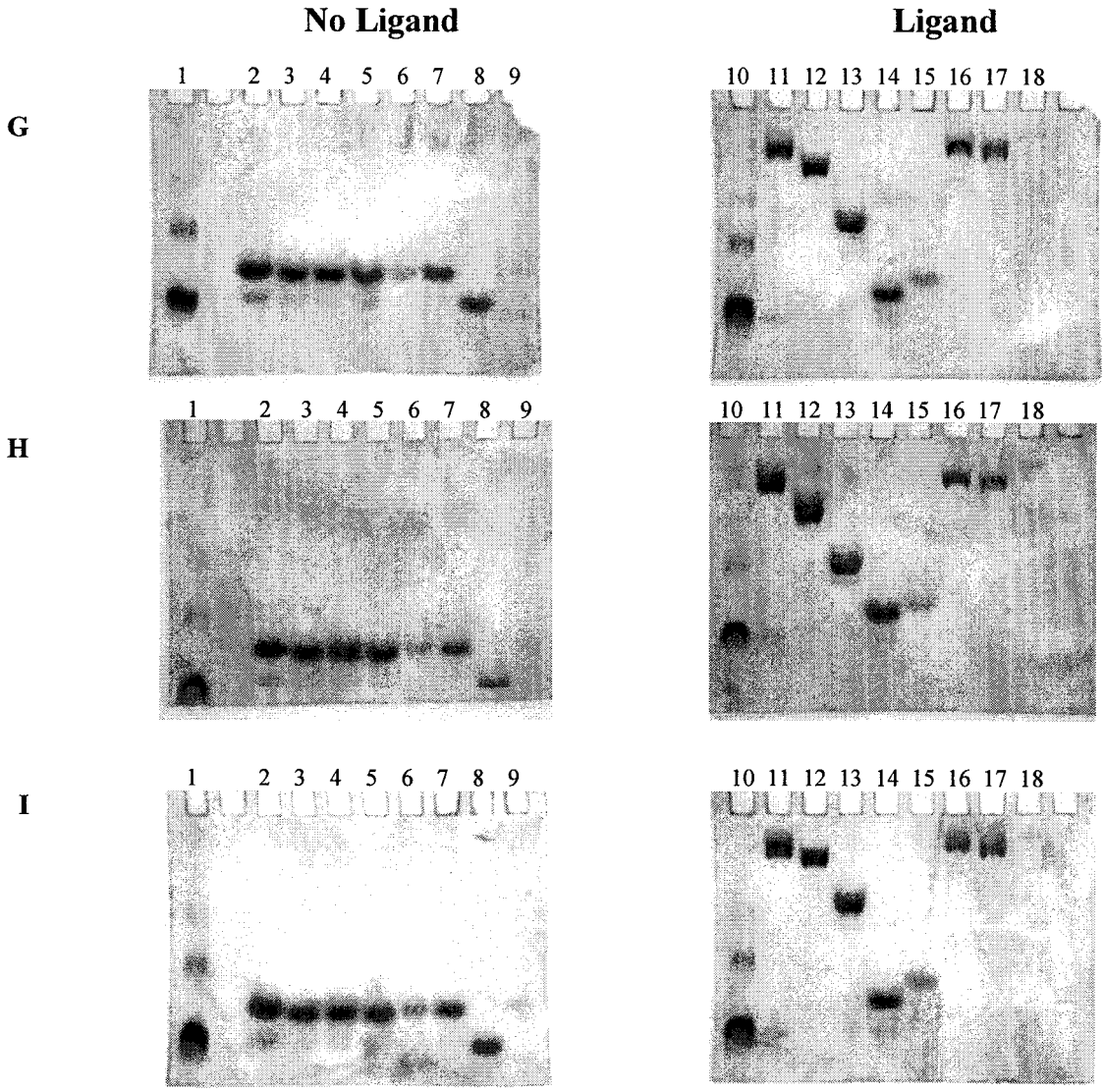


Figure 3.12: Quantitative AGE of wild type *Ct*CBM6 and the mutants in gels containing different concentrations of oat spelt xylan.
Wild type *Ct*CBM6, the mutants (Y34A, W92M, N120A, Y40M, R72M and Y112M) and BSA (non-binding control) were subjected to AGE, using a 7.5 % polyacrylamide gel in a Tris/glycine buffer system, pH 8.3, in the presence of different concentrations (final concentrations) of soluble oat spelt xylan, as described in Section 2.2.4.1. The gels (with and without ligand) were electrophoresed for 3 h at 10 mA/gel. Proteins were visualised by staining with Coomassie Blue and destained (Section 2.2.3.1.1). (G) 0.06 mg/ml: No ligand: Lane 1: 5 µg BSA; lane 2: 10 µl wild type *Ct*CBM6; lane 3: 10 µl Y95F; lane 4: 10 µl Y34A; lane 5: 10 µl N120A; lane 6: 10 µl W92M; lane 7: 10 µl Y40M; lane 8: 10 µl R72M; lane 9: 10 µl Y112M. Ligand: lane 10: 5 µg BSA; lane 11: 10 µl wild type *Ct*CBM6; lane 12: 10 µl Y95F; lane 13: 10 µl Y34A; lane 14: 10 µl N120A; lane 15: 10 µl W92M; lane 16: 10 µl Y40M; lane 17: 10 µl R72M; lane 18: 10 µl Y112M. (H) 0.1 mg/ml: No ligand: Lane 1: 5 µg BSA; lane 2: 10 µl wild type *Ct*CBM6; lane 3: 10 µl Y95F; lane 4: 10 µl Y34A; lane 5: 10 µl N120A; lane 6: 10 µl W92M; lane 7: 10 µl Y40M; lane 8: 10 µl R72M; lane 9: 10 µl Y112M. Ligand: lane 10: 5 µg BSA; lane 11: 10 µl wild type *Ct*CBM6; lane 12: 10 µl Y95F; lane 13: 10 µl Y34A; lane 14: 10 µl N120A; lane 15: 10 µl W92M; lane 16: 10 µl Y40M; lane 17: 10 µl R72M; lane 18: 10 µl Y112M. (I) 0.2 mg/ml: No ligand: Lane 1: 5 µg BSA; lane 2: 10 µl wild type *Ct*CBM6; lane 3: 10 µl Y95F; lane 4: 10 µl Y34A; lane 5: 10 µl N120A; lane 6: 10 µl W92M; lane 7: 10 µl Y40M; lane 8: 10 µl R72M; lane 9: 10 µl Y112M. Ligand: lane 10: 5 µg BSA; lane 11: 10 µl wild type *Ct*CBM6; lane 12: 10 µl Y95F; lane 13: 10 µl Y34A; lane 14: 10 µl N120A; lane 15: 10 µl W92M; lane 16: 10 µl Y40M; lane 17: 10 µl R72M; lane 18: 10 µl Y112M.

J

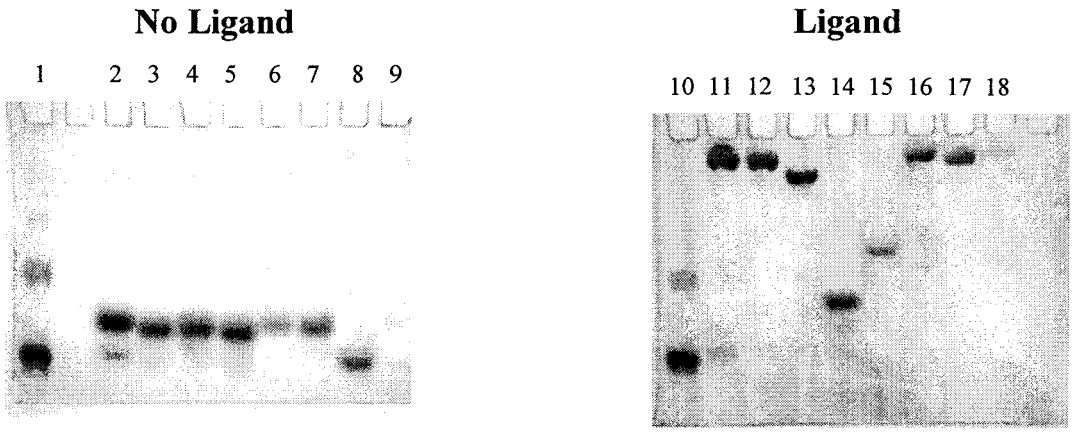


Figure 3.13: Quantitative AGE of wild type *Ct*CBM6 and the mutants in gels containing different concentrations of oat spelt xylan.
Wild type *Ct*CBM6, the mutants (Y34A, W92M, N120A, Y40M, R72M and Y112M) and BSA (non-binding control) were subjected to AGE, using a 7.5 % polyacrylamide gel in a Tris/glycine buffer system, pH 8.3, in the presence of different concentrations (final concentrations) of soluble oat spelt xylan, as described in Section 2.2.4.1. The gels (with and without ligand) were electrophoresed for 3 h at 10 mA/gel. Proteins were visualised by staining with Coomassie Blue and destained (Section 2.2.3.1.1). (J) 1.0 mg/ml: No ligand: Lane 1: 5 μg BSA; lane 2: 10 μl wild type *Ct*CBM6; lane 3: 10 μl Y95F; lane 4: 10 μl Y34A; lane 5: 10 μl N120A; lane 6: 10 μl W92M; lane 7: 10 μl Y40M; lane 8: 10 μl R72M; lane 9: 10 μl Y112M. Ligand: lane 10: 5 μg BSA; lane 11: 10 μl wild type *Ct*CBM6; lane 12: 10 μl Y95F; lane 13: 10 μl Y34A; lane 14: 10 μl N120A; lane 15: 10 μl W92M; lane 16: 10 μl Y40M; lane 17: 10 μl R72M; lane 18: 10 μl Y112M.

Variant of CBM6	K_a (%)	K_d (%)
WT	436.68	0.00229
Y34A	52.63	0.019
N120A	3.32	0.301
W92M	8.40	0.119
Y40M	512.82	0.00195
R72M	473.93	0.00211
Y112M	520.83	0.00192

Table 3.3: Affinity of *Ct*CBM6 (wild type) and mutants for oat spelt xylan using quantitative AGE.
K_a (affinity association constant) and K_d (dissociation constant) for *Ct*CBM6 (wild type) and mtants (Y34A, W92M, N120A (Cleft A); Y40M, R72M, Y112M (Cleft B)). Experiments were carried out as described in Section 2.2.4.1.

The migration distance of the mutants Y34A, W92M and N120A (Figures 3.10-3.13) were not retarded greatly by the presence of xylan in the gel compared to the wild type protein (Figures 3.10-3.13). Determination of the K_a values showed that Y34A, W92M and N120A displayed reduced affinity for oat spelt xylan, compared to the wild type CBM, and therefore Tyr-34, Trp-92 and Asn-120 (located in Cleft A) play a pivotal role in the interaction of *Ct*CBM6 with its ligand, either by interacting directly with xylan or in maintaining the structure of the binding cleft. In contrast, mutants Y40M, R72M and Y112M showed a broadly similar affinity for oat spelt xylan to the wild type CBM, as indicated by their K_a values (Table 3.3). It can therefore be concluded that Tyr-40, Arg-72 and Tyr-112, which are located in Cleft B, do not play a role in the interaction of *Ct*CBM6 with its ligand.

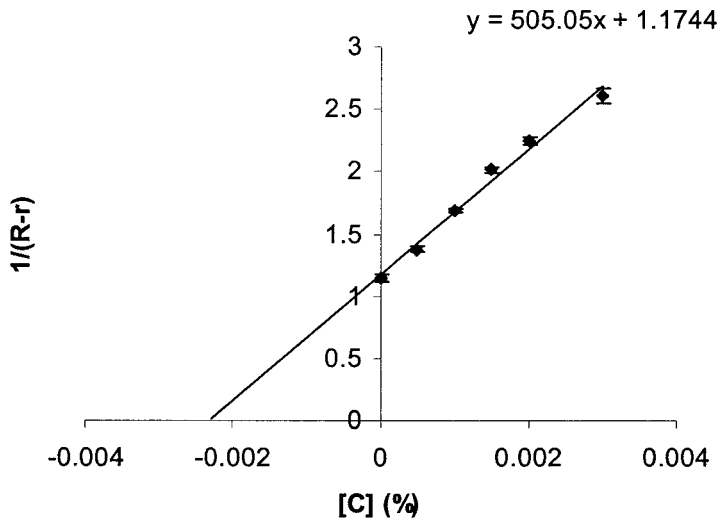


Figure 3.14: Reciprocal plot of relative mobility of *Ct*CBM6 (wild type) against oat spelt xylan concentration.
The relative mobility (R-r) was calculated as described in section 3.3.2. *Ct*CBM6 was retarded in the presence of increasing concentrations of oat spelt xylan, while the control protein (BSA) was not affected by this ligand. The dissociation constant (K_d) can be estimated from the equation of the line; the value x when $y = 0$ is the negative value of K_d. These data were performed in triplicate with error bars representing the standard deviation from the mean.

To summarise, qualitatively Y34A, W92M and N120A bound to xylan (Table 3.2), however, quantitative AGE (Table 3.3) revealed that the affinities of these mutants for oat spelt xylan were 8-fold, 52-fold and 132-fold, respectively, less than that of the wild type, indicating that these mutants had significantly reduced xylan binding capacity. These data led to the conclusion that Tyr-34, Trp-92 and Asn-120 are important in ligand binding and that the location of the ligand-binding site in *Ct*CBM6 is Cleft A.

3.4 Thermodynamics of xylooligosaccharides and xylan binding to *Ct*CBM6

ITC was used to quantify the affinity of *Ct*CBM6 (wild type only) for various soluble ligands, and to determine the thermodynamic parameters for these binding events. Briefly, the K_a of *Ct*CBM6 for its ligands (K_a) and the change in enthalpy ($\Delta H°$) were determined experimentally from each ITC curve and used to calculate the change in Gibbs free energy ($\Delta G°$) and $T\Delta S°$ (change in entropy) for each ligand-binding event as described in section 2.2.4.

The data for typical ITC curves generated when *Ct*CBM6 was titrated with a series of xylooligosaccharides, xylan and cellooligosaccharides are shown in figures 3.15-3.18 and the full data set is presented in Table 3.4. Binding to each oligosaccharide may be fitted by a one-binding site model, supporting the view that the protein contains only a single ligand-binding site. The results show that *Ct*CBM6 binds preferentially to xylooligosaccharides and xylan, interacts weakly with cellooligosaccharides but does not interact with xylose.

The affinity of *Ct*CBM6 for xylooligosaccharides was shown to increase up to xylopentaose. The K_a of the CBM for xylopentaose is ~ 50 and ~ 100 times higher than for xylobiose and cellohexaose, respectively. As the K_a value did not change significantly between xylopentaose and xylohexaose, it is likely that the binding site of *Ct*CBM6 can accommodate five xylose residues. *Ct*CBM6 showed higher affinity for xylopentaose and xylohexaose than for oat spelt xylan as judged by ITC, which could reflect the lower purity of the polysaccharide. The affinity of the protein for highly substituted arabinoxylans (rye and wheat) or poorly substituted xylans (oat spelt and birchwood) is broadly similar. These data indicate that *Ct*CBM6 can accommodate highly decorated xylans. The stoichiometry of binding for all oligosaccharides at saturation was 1:1, indicating that there is only one ligand-binding site per protein molecule. These data reinforces the view that CBM6 from *C. thermocellum* Xyn11A primarily binds to xylan.

In all binding events there was a net decrease in entropy, $T\Delta S°$, and the enthalpy, $\Delta H°$, values were large, in a range of – 6.6 to – 19.8 kcal mol^{-1}. These thermodynamic parameters show that the interactions between *Ct*CBM6 and its ligands were enthalpically driven and therefore exothermic.

To summarise, the data (Table 3.4) shows that *Ct*CBM6 binds preferentially to xylooligosaccharides and xylan, interacts only weakly with cellooligosaccharides, and exhibits no detectable affinity for xylose. There is a progressive increase in ligand affinity up to xylopentaose, suggesting that the ligand-binding site can accommodate upto five xylose residues. The thermodynamic parameters (Table 3.4) indicate that the binding interaction for all ligands tested was mainly driven by a favourable enthalpy. It can be concluded, therefore, that *Ct*CBM6 primarily interacts with xylan of varying degrees of substitution.

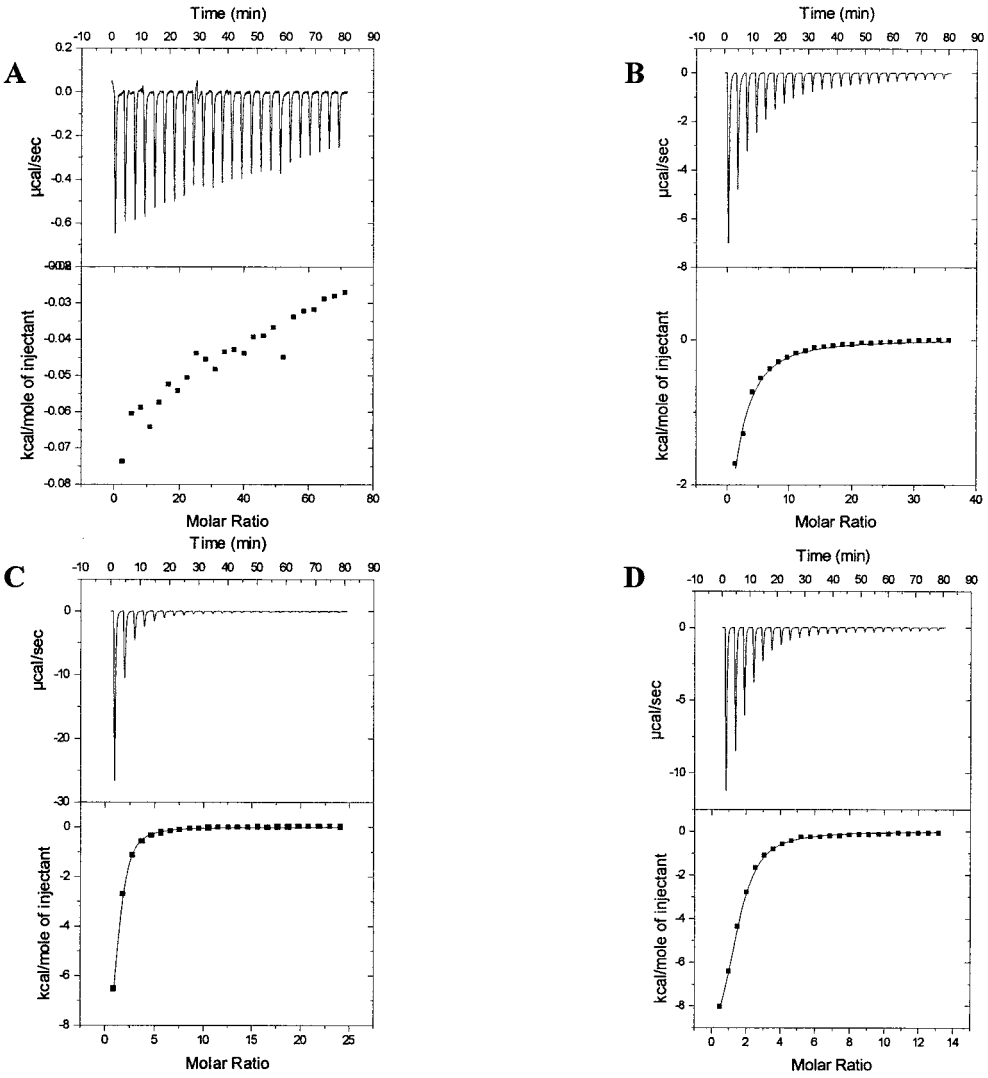


Figure 3.15: Isothermal titration calorimetry (ITC) analysis of the binding of xylooligosaccharides to *Ct*CBM6.

*Ct*CBM6 at a concentration of 114.0 µM was titrated with A. Xylose (30 mM) B. Xylobiose (15 mM) C. Xylotriose (15 mM) D. Xylotetraose (5 mM). The top half of each panel shows the calorimetric titration of *Ct*CBM6 with ligand, and the lower half displays the integrated injection heats from the upper panel, corrected for control dilution heats. The solid line is the curve of best fit that was used to derive parameters *n* (stoichiometry), K_a and ΔH°. ΔG° was calculated from the standard thermodynamic equation. Molar ratio is the ratio of ligand to protein. All titrations were performed in 50 mM sodium phosphate buffer, pH 7.0, at 25 °C, as detailed in Section 2.2.4.2.

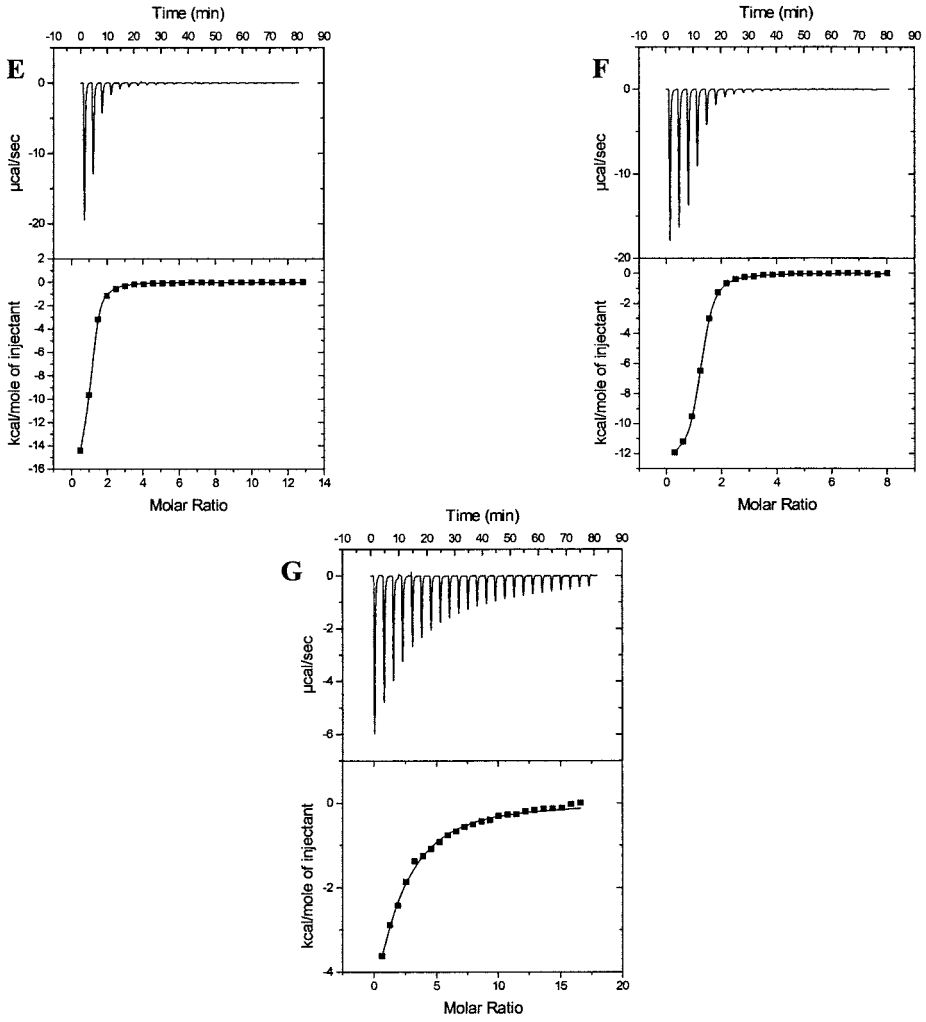


Figure 3.16: Isothermal titration calorimetry (ITC) analysis of the binding of xylooligosaccharides to *Ct*CBM6.

*Ct*CBM6 at a concentration of 114.0 µM was titrated with E. Xylopentaose (5 mM, 25 °C) F. Xylohexaose (5 mM, 25 °C) and G. Xylopentaose (5 mM, 60 °C). The top half of each panel shows the calorimetric titration of *Ct*CBM6 with ligand, and the lower half displays the integrated injection heats from the upper panel, corrected for control dilution heats. The solid line is the curve of best fit that was used to derive parameters *n* (stoichiometry), K_a and ΔH°. ΔG° was calculated from the standard thermodynamic equation. Molar ratio is the ratio of ligand to protein. All titrations were performed in 50 mM sodium phosphate buffer, pH 7.0, as detailed in Section 2.2.4.2.

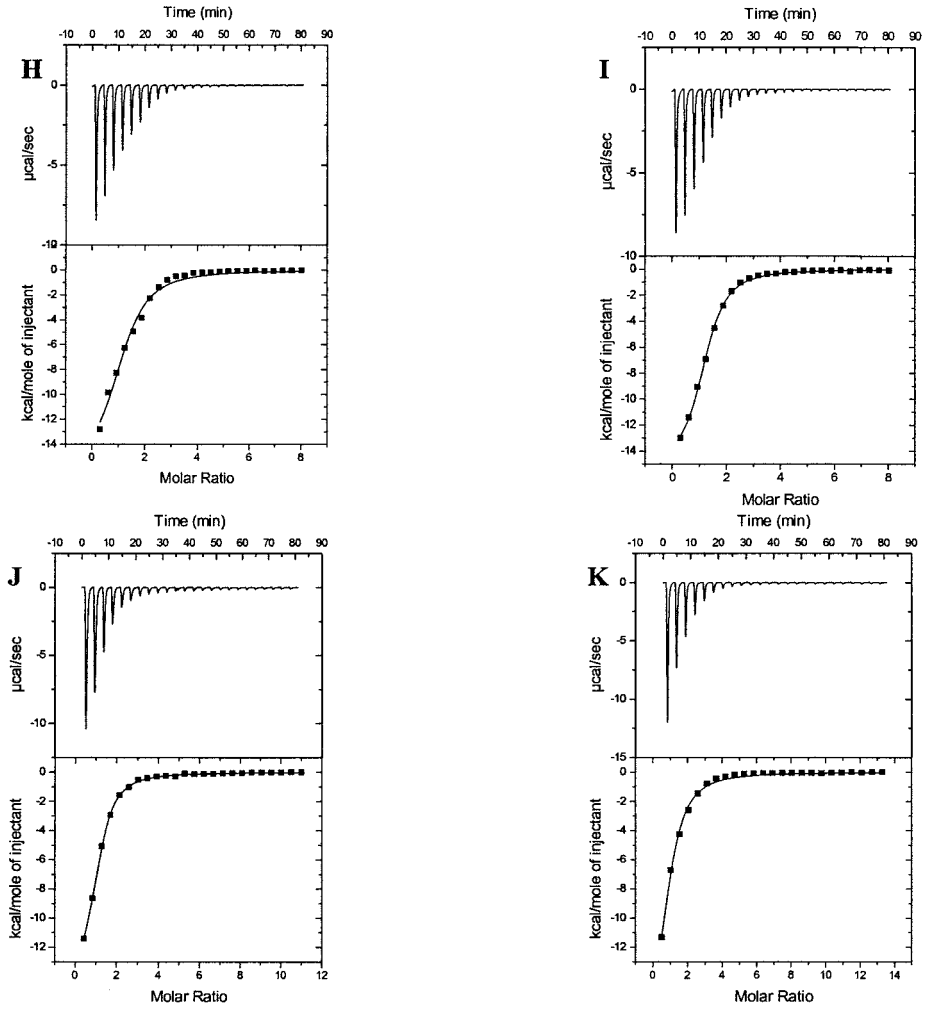


Figure 3.17: Isothermal titration calorimetry (ITC) analysis of the binding of xylan and arabinoxylan to *Ct*CBM6.

*Ct*CBM6 at a concentration of 114.0 µM was titrated with H. Oat spelt xylan I. Birchwood xylan J. Rye arabinoxylan K. Wheat arabinoxylan. The top half of each panel shows the calorimetric titration of *Ct*CBM6 with ligand, and the lower half displays the integrated injection heats from the upper panel, corrected for control dilution heats. The solid line is the curve of best fit that was used to derive parameters *n* (stoichiometry), K_a and ΔH°. ΔG° was calculated from the standard thermodynamic equation. Molar ratio is the ratio of ligand to protein. All titrations were performed in 50 mM sodium phosphate buffer, pH 7.0, at 25 °C, as detailed in Section 2.2.4.2.

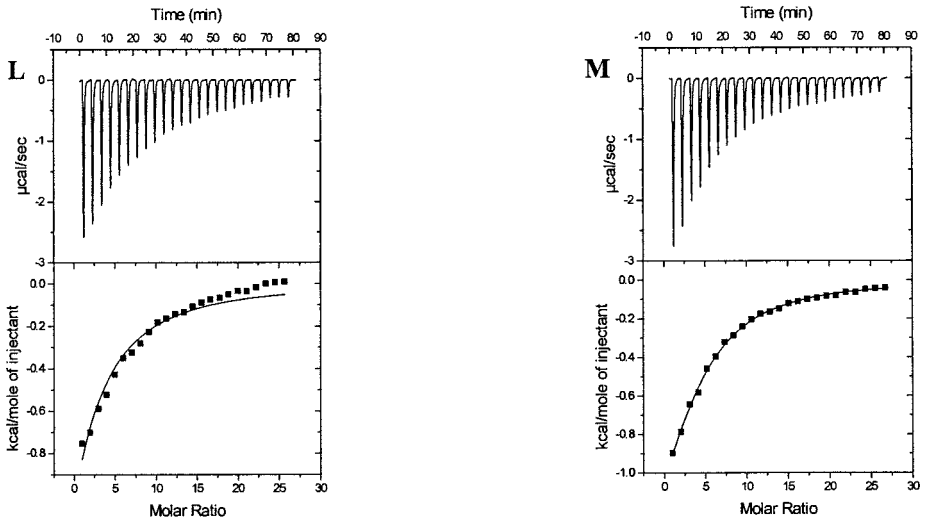


Figure 3.18: Isothermal titration calorimetry (ITC) analysis of the binding of cellooligosaccharides to *Ct*CBM6.

*Ct*CBM6 at a concentration of 114.0 µM was titrated with L. Cellohexaose M. Cellopentaose. The top half of each panel shows the calorimetric titration of *Ct*CBM6 with ligand, and the lower half displays the integrated injection heats from the upper panel, corrected for control dilution heats. The solid line is the curve of best fit that was used to derive parameters *n* (stoichiometry), K_a and ΔH°. ΔG° was calculated from the standard thermodynamic equation. Molar ratio is the ratio of ligand to protein. All titrations were performed in 50 mM sodium phosphate buffer, pH 7.0, at 25 °C, as detailed in Section 2.2.4.2.

Ligand	K_a x 10^{-3} (M^{-1})	ΔG (kcal mol^{-1})	ΔH (kcal mol^{-1})	$T\Delta S$ (kcal mol^{-1})	n^a
Xylose	ND[b]	ND	ND	ND	ND
Xylobiose	3.7 ± 0.4[c]	-4.9 ± 0.1	-9.1 ± 1.6	-4.2 ± 1.7	1.16
Xylotriose	25.2 ± 5.0	-6.0 ± 0.1	-13.0 ± 1.6	-7.0 ± 1.6	0.96
Xylotetraose	43.4 ± 10.1	-6.3 ± 0.1	-11.1 ± 0.1	-4.8 ± 0.2	1.33
Xylopentaose	175.0 ± 24.2	-7.1 ± 0.1	-13.5 ± 0.1	-6.6 ± 0.1	1.07
Xylohexaose	189.0 ± 21.0	-7.2 ± 0.1	-13.2 1.1	-6.0 ± 1.1	1.18
Oat spelt xylan	58.7 ± 9.7	-6.5 ± 0.3	-16.1 ± 1.0	-9.6 ± 2.6	1.14
Birchwood xylan	102.0 ± 5.6	-6.8 ± 0.3	-15.2 ± 0.2	-8.4 ± 2.3	1.16
Wheat arabinoxylan	42.6 ± 4.7	-6.3 ± 0.4	-19.8 ± 1.5	-13.5 ± 3.0	0.82
Rye arabinoxylan	70.5 ± 2.5	-6.6 ± 0.7	-15.0 ± 0.2	-8.4 ± 0.1	0.98
Cellohexaose	1.5 ± 0.4	-4.3 ± 0.16	-6.6 ± 1.7	-2.3 ± 1.5	1.2
Cellopentaose	3.07 ± 0	-4.8 ± 0.18	-2.12 ± 1.8	-2.6 ± 1.6	1.3

Table 3.4: Affinity values and thermodynamic parameters of *Ct*CBM6 against oligosaccharides and polysaccharides as determined by ITC.

Three individual titrations with each ligand were carried out as detailed in Section 2.2.4.2. [a] number of binding sites on the protein, [b] no detectable binding, [c] ± standard deviation of three titration experiments.

Investigation of the interactions between *Clostridium thermocellum* Xyn11A CBM6 and its ligands.

4 Discussion

4.1 Introduction

The primary aim of work undertaken in this chapter was to investigate the ligand specificity of *Clostridium thermocellum* Xyn11A CBM6 (*Ct*CBM6), and to relate this data to the three dimensional (3D) structure of the domain, which was solved by Dr M. Czjzek at CNRS, Marseille. Affinity gel electrophoresis (AGE; Section 2.2.4.1) and isothermal titration calorimetry (ITC; Section 2.2.4.2) were used to determine the ligand specificity of *Ct*CBM6, and to calculate its affinity for a variety of soluble polysaccharides and the thermodynamic forces driving these interactions. Structural studies showed that ligand-binding in *Ct*CBM6 contains two distinct clefts (designated Cleft A and Cleft B) that could comprise the ligand-binding site. To identify the location of the ligand-binding site(s) in *Ct*CBM6, a mutagenesis strategy was employed to support NMR data.

4.2 CBMs with xylan affinity from *Clostridia* enzymes

The majority of glycoside hydrolases (GHs) comprise, in addition to the catalytic modules, one or more ancillary non-catalytic modules, which frequently bind polysaccharides. These hydrolases are referred to as modular enzymes (Warren, 1996). Most modular *Clostridia* plant cell wall (PCW) hydrolases contain CBMs and those that are derived from GH family 10 and 11 xylanases display affinity for xylan. For example, *C. stercorarium* contains three xylan-binding family 6 CBMs (*Cs*CBM6-1, *Cs*CBM6-2, and *Cs*CBM6-3) (Sun *et al.*, 1998; Boraston *et al.*, 2003); *C. thermocellum* contains two CBMs from family 22 (*Ct*CBM22a and *Ct*CBM22b) (Charnock *et al.*, 2000; Xie *et al.*, 2001b) and also a family 6 CBM (*Ct*CBM6) (Fernandes *et al.*, 1999), which has been the focus of this chapter. It is presently unclear why two or more CBMs should be present in the same xylan-degrading enzymes. One possibility is that the modules function in synergy to increase the affinity of the enzyme for its soluble substrate. Linder *et al.* (1996)

demonstrated that a recombinant protein, composed of two CBM1s, showed a higher affinity for cellulose than the sum of the affinities of the individual modules. Xyn11A CBM2b-1 and CBM2b-2 from *Cellulomonas fimi* (*Cf*Xyn11A) have been shown to act synergistically to increase the affinity of the full-length enzyme for xylan and cellulose compared to the individual modules. Indeed, discrete CBM2bs do not appear to interact with cellulose (Bolam *et al.*, 1999).

Another possibility for the presence of two CBMs in a single enzyme is that the individual modules bind to different polysaccharides, as proposed by Gill *et al.* (1999). It is possible to envisage that at any one time during the degradation process, different polysaccharides would be exposed to the aqueous environment. Thus, by having two CBMs with different ligand specificity, these enzymes are likely to interact with the PCW, regardless of its degradation state.

4.3 Three dimensional structure of *Ct*CBM6

The CBMs described to date have a "jelly-roll" fold composed almost exclusively of β-strands. The structures of CBMs from 22 of the 39 families have been determined and fall into one of three types of modules. For example, CBMs that bind to crystalline cellulose, known as Type A modules, contain a planar hydrophobic ligand-binding surface (Xu *et al.*, 1995; Raghothama *et al.*, 2000). By contrast CBMs that interact with individual polysaccharide chains, known as Type B modules, accommodate their target ligands in a cleft of varying depth (Charnock *et al.*, 2000; Czjzek *et al.*, 2001; Boraston *et al.*, 2002a). The final type of module, Type C, bind to the ends of glycoside chains. Studies have shown that generally both Type A and Type B CBMs interact with five or six saccharide units via hydrophobic stacking interactions between alternate sugar rings and aromatic amino acids, although hydrogen bonds also play an important role in carbohydrate recognition by Type B modules (Xie *et al.*, 2001a; 2001b; Pell *et al.*, 2003; Henshaw *et al.*, 2004).

All structural work on *Ct*CBM6 was carried out by Dr M. Czjzek at CNRS, Marseille. The structure of *Ct*CBM6 forms a classic lectin-like β-jelly roll (Figure 3.2), predominantly consisting of five antiparallel β-strands on one face and four

antiparallel β-strands on the other face (Czjzek *et al.*, 2001). This type of fold in CBMs is common with the topology of the ligand-binding site reflecting the nature of the target ligand. For example, the crystal structure of CBM22b was resolved by Charnock *et al.* (2000) and its overall topology is a typical β-jelly roll consisting of two four β-stranded sheets. The ligand-binding site in CBM22b is a shallow groove, along the surface of the jelly roll structure. The structure of *Ct*CBM6 contains aromatic residues on the surface of two putative ligand-binding sites, which are common in CBMs (Charnock *et al.*, 2000; Simpson *et al.*, 2000; Xie *et al.*, 2001b). Similar solvent-exposed aromatic amino acids have been shown to be directly involved in ligand-binding in CBMs from different families (Din *et al.*, 1994; Linder *et al.*, 1995; Nagy *et al.*, 1998; Simpson & Barras, 1999; Kormos *et al.*, 2000; Ponyi *et al.*, 2000). Also, there are polar residues flanking the aromatic amino acids in cleft A and B of *Ct*CBM6. Hydrophobic stacking and hydrogen bonding interactions between ligands and surface aromatic and polar amino acids are thought to be the main forces involved in the association of protein with carbohydrates. *Ct*CBM6 displays a small shallow cleft, lined by two aromatic residues Tyr-34 and Trp-92 (Figure 3.2; Cleft A). There is also a second shallow cleft on the concave surface of *Ct*CBM6, lined by two aromatic residues Tyr-40 and Tyr-112 (Figure 3.2; Cleft B). Initially, cleft B was identified as the primary binding cleft in *Ct*CBM6, as this resembled the ligand-binding site in CBM4 and CBM22. NMR and mutational analysis was then carried out on both clefts to determine the primary binding cleft in *Ct*CBM6.

4.4 Identification of the ligand-binding site of *Ct*CBM6 by NMR

Structural and mutagenesis studies have indicated that the binding of CBMs from families 2b, 4b, 13, 15, and 22b (Clarke *et al.*, 1996; Kittur *et al.*, 2003; Nordberg *et al.*, 1997; Dupont *et al.*, 1998; Fujimoto *et al.*, 2002; Szabó *et al.*, 2001; Pell *et al.*, 2003; Charnock *et al.*, 2000; Xie *et al.*, 2001a) to xylan is mediated primarily by a cleft containing a strip of solvent exposed aromatic residues (usually tryptophans and tyrosines), which are thought to form hydrophobic stacking interactions with the xylopyranose rings of the xylose moieties of one xylan chain (Charnock *et al.*, 2000; Simpson *et al.*, 2000; Xie *et al.*, 2001a). Hydrogen-bonding between polar residues flanking this aromatic strip and hydroxyl groups

in the sugar chain with which ring stacking occurs, as well as with adjacent chains, also appear to contribute to the binding event (Xu *et al.*, 1995; Johnson *et al.*, 1996; Simpson *et al.*, 1999). Two solvent exposed aromatic residues (Tyr-34 and Trp-92; Cleft A), which are not invariant within family 6 CBMs, form a shallow groove at the surface of *Ct*CBM6 and could potentially be part of the ligand-binding site. Thus, these residues are predicted to possibly be key components of the ligand-binding site in *Ct*CBM6. However, structural analysis revealed a second putative binding cleft (Cleft B), which also had two surface exposed aromatic residues that are more strongly conserved in family 6 CBMs (Tyr-40, replaced by a tryptophan in a number of family members, and Tyr-112, replaced by phenylalanine, threonine or valine in a small number of families; Figure 3.3). Cleft B in *Ct*CBM6 showed similarities to the binding clefts observed in families 4 (Johnson *et al.*, 1996) and 22 (Charnock *et al.*, 2000), thus, it was possible that these residues could also be key components of the ligand-binding site in *Ct*CBM6. In order to investigate the precise location on *Ct*CBM6 where xylohexaose, and thus by inference xylan, was binding, the protein was titrated with xylohexaose and 2D ^{1}H and ^{15}N heteronuclear single quantum coherence (HSQC) NMR spectra recorded over the course of the titration (NMR studies were carried out by Dr M. Czjzek, CNRS, Marseille). The data showed that the largest changes in chemical shifts were those of the side chains of Tyr-34 and Trp-92 (Czjzek *et al.*, 2001), located in cleft A, indicating that they are part of the ligand-binding cleft. Other residues also showed significant changes in their chemical shifts and were found to be close to cleft A. None of the residues situated in cleft B were affected by the titration experiment, indicating that this region does not constitute the ligand-binding cleft in *Ct*CBM6. Surrounding the two aromatic rings in the binding site of cleft A are hydrophilic residues, including Asn-120. In sugar-binding proteins such as lectins, hydrogen-bonding between the protein and sugar most commonly occurs via these types of side chains (Weis and Drickhamer, 1996), i.e. through interactions of sugars with amide and carboxylate groups. The residue Asn-120 is orientated such that its side chain group is able to form hydrogen bonds to a polysaccharide stacking between the exposed tryptophan and tyrosine and, correspondingly, the largest chemical shift perturbation on ligand-binding was observed in this side chain amide (Czjzek *et al.*, 2001). Therefore, this residue is also likely to be involved in ligand-binding

through hydrogen-bonding. In order to demonstrate the location of the ligand-binding site in *Ct*CBM6, mutants of the likely key residues in cleft A and cleft B were constructed and their capacity to bind xylan was assessed by AGE.

4.5 Ligand specificity of *Ct*CBM6

CBM family 6 contains modules from 54 enzymes (CAZy). Members of family 6 that bind to cellulose, xylan, mixed β-1,3-1,4-glucan and β-1,3-glucan have been described, and putative family 6 CBMs are also present in an α-agarase and an α-1,6-mannanase (CAZy). The first structure determination for a family 6 CBM, that from *Ct*CBM6 (Czjzek *et al.*, 2001), revealed a β-sheet fold similar to those of CBMs from families 4 and 22, but with the binding site in a different location. Family 6 appeared to offer an opportunity for determining how a conserved fold may present different binding site topologies and modes of interaction to create different carbohydrate binding specificities.

*Ct*CBM6 has been demonstrated from previous work (Fernandes *et al.*, 1999) and from this study to be a functional CBM that bound preferemtially to xylan. Although *Ct*CBM6 bound to the four different forms of xylan, it also interacted to a small extent with other polysaccharides such as β-1,4-glucan-linked carboxymethyl-, hydroxyethyl- and methyl- cellulose and the mixture of β-1,3-β-1,4-glycosidic linked barley β-glucan. Thus it can be inferred that the binding site can mainly accommodate the β-1,4-linked xylose chain of the hemicellulose, but has a small degree of flexibility to allow the weak binding of other polysaccharides. However, it is likely that a stable protein-ligand complex is observed when the protein interacts with the xylan backbone. From the structure of this module, *Ct*CBM6 has a shallow groove, thus there is a space for the side chains of substituted xylans to bind. This is seen for other modules that have had their structure determined. For example, *Cf*Xyn11A CBM2b-1 and *Ct*Xyn10B CBM22b, which also bind to soluble and insoluble xylans and substituted xylans, have a relatively open cleft, and thus there is a space for the side chains (Simpson *et al.*, 1999; Charnock *et al.*, 2000). CBM13 from *Streptomyces olivaceoviridis* Xyn10A also interacts with decorated xylans and the 3D structure in complex with α-L-arabinofuranosyl- and 4-*O*-methyl-α-D-glucuronosyl-

xylooligosaccharides demonstrates the accommodation of the side chains in the cleft (Fujimoto *et al.*, 2004). CBM15 from *Cellvibrio japonicus* (formerly *Pseudomonas cellulosa*) Xyn10C also binds xylan, including decorated xylans and xylooligosaccharides. The 3D structure reveals a β-jelly roll structure displaying an extended groove that runs along the concave face of the protein, forming the binding site for target ligands (Szabó *et al.*, 2001). Within this cleft, two tryptophan residues form hydrophobic stacking interactions with two xylopyranose units reflecting the approximate 3-fold helical conformation of bound xylan. The location of the polysaccharide solvent-exposed 2' and 3' hydroxyls reveals how these xylan-binding CBMs can accommodate both decorated and naked xylans (Szabó *et al.*, 2001).

Hydrophobic stacking interactions and hydrogen bonding are the main forces that mediate the association of CBMs with carbohydrates (Din *et al.*, 1994; Linder *et al.*, 1995; Bray *et al.*, 1996; Matteinen *et al.*, 1997; Nagy *et al.*, 1998; Ponyi *et al.*, 2000). The data presented in this chapter provides further evidence that aromatic residues in the binding cleft are important in ligand binding. Amino acids located on the surface of the two clefts (A and B) of *Ct*CBM6 were mutated, and the ligand binding capacity of the resultant proteins was assessed. Mutations to Tyr-40, Arg-72, and Tyr-112, located in Cleft B, did not result in a significant decrease in affinity of the protein for xylan (Table 3.3). In contrast, the mutants W92M, Y34A, and N120A exhibited greatly reduced affinity for oat spelt xylan. The location of Tyr-34, Trp-92 and Asn-120 in Cleft A indicates that this region comprises the ligand-binding site of *Ct*CBM6. The possible explanation for the greatly reduced loss of binding by these mutations is that Tyr-34, Trp-92 and Asn-120 are directly involved in ligand binding.

In addition to the hydrophobic interactions, surface polar residues are thought to hydrogen bond to the hydroxyl groups of the sugar molecules in the target carbohydrate ligands of CBMs (Xu *et al.*, 1995; Johnson *et al.*, 1996; Tormo *et al.*, 1996). For example, Asn-120, which is located on the binding face of *Ct*CBM6, plays an important role in the binding of the protein to its ligands. The hydrogen bonding interactions contributed by Asn-120 was critical, as the mutant

N120A showed a significant reduction in affinity for its target ligands when this residue was mutated (Table 3.3). This finding supports the proposal that polar amino acids in the binding cleft are directly involved in ligand binding, by forming hydrogen bonds with the sugar molecules. It is possible, however, that the loss of affinity was caused by structural changes in the binding site, resulting in the key carbohydrate-binding residues not being in the proper position to interact with their target ligand.

The observation that the backbones of all the ligands, which *Ct*CBM6 binds to comprise β-1,4-linked xylosyl residues, indicates that the β-1,4-linked polymer is the ligand that *Ct*CBM6 recognises. The fact that *Ct*CBM6 interacts with both xylan and cellulose, albeit only weakly, suggests that the surface of this protein can accommodate both twisted and flat ligands, as xylan and cellulose comprise 3-fold helical and planar structures, respectively. Also it is interesting to note that other CBMs, which display affinity for both soluble and insoluble xylan and cellulose, have recently been reported (Sun *et al.*, 1998; Simpson *et al.*, 1999; Abou Hachen *et al.*, 2000; Boraston *et al.*, 2003; Henshaw *et al.*, 2004). An explanation for this phenomenon is that the binding site contains residues, located in different orientations that bind to distinct ligands, or the orientation of surface residues is flexible, and a conformational change occurs when the protein associates with its ligand. Another possibility is that the protein binds to terminal sugar residues of xylan and cellulose and that these interactions do not involve the C6OH of glucose. Such a phenomenon has been observed in both a CBM9 (Notenboom *et al.*, 2001) and in the recently characterised CBM6 from *Cellvibrio mixtus* endoglucanase 5A (Pires *et al.*, 2004).

Although xylose and glucose are chemically very similar, the only difference being that xylose lacks the CH_2OH group present at the C5 in glucose, the structures of the polysaccharides generated from β-1,4-linked glucose and xylose molecules, respectively, are strikingly different. In cellulose chains, cellobiose is the basic repeating unit, which causes the chain to adopt a planar conformation i.e. the chains pack closely together in layered sheets (Tormo *et al.*, 1996). Xylan, due to the missing CH_2OH group, does not pack in such an ordered way. Xylotriose is

the basic repeating unit in xylan and causes the chains to adopt a helical, rather than flat conformation (Atkins, 1992). This difference in secondary structure of the polysaccharides appears to explain how the orientation of the exposed aromatic residues leads to the difference in ligand specificity between different families of CBMs. CBMs that recognise planar cellulose chains use coplanar aromatic residues, whereas CBMs that recognise the twisted xylan chain use aromatic residues arranged perpendicular to each other (Simpson *et al.*, 2000).

4.6 Interaction mechanism between *Ct*CBM6 and the ligand

*Ct*CBM6 associates with both soluble and insoluble polysaccharides (Fernandes *et al.*, 1999; this study). The solved 3D structure of *Ct*CBM6, the first for a family 6 CBM, possesses a shallow groove/cleft that can accommodate the flexible chain of soluble ligands and also the rigid structure of insoluble polymers (Czjzek *et al.*, 2001).

The thermodynamics of ligand binding to *Ct*CBM6 were investigated by ITC. Two or three independent titration experiments were performed with each ligand. Stoichiometries, enthalpies and association constants were determined by fitting the corrected data to a 1:1 bimolecular interaction model. The data (Table 3.4) indicate that binding is driven primarily by enthalpic forces ($\Delta H°$), while the entropic contribution ($T\Delta S°$) to binding is unfavourable. These results are consistent with other soluble carbohydrate/protein interactions, which are also driven by enthalpic forces and have $\Delta H°$ values ranging from 4-25 kcal/mol (Creagh *et al.*, 1996). It has been argued that enthalpically driven binding reflects the dominance of hydrogen bonds and van der Waals forces in the interaction (Tomme *et al.*, 1996), a view supported by the availability of residues capable of hydrogen bonding with the sugar in *Ct*CBM6 (Figure 3.3).

ITC was used to quantify the interaction of *Ct*CBM6 with its target ligands. Representative titration's that can be deconvoluted to give thermodynamic and binding parameters are displayed in Figures 3.10-3.13 and the full data set presented in Table 3.4. The data presented in Table 3.4 reveal that the binding of *Ct*CBM6 to its ligands is associated with negative enthalpy and entropy. The

thermodynamics of the interaction of *Ct*CBM6 with xylose-based ligands is typical of the binding of proteins to soluble saccharides. Binding appears to be driven mainly by enthalpic forces, partly compensated by a negative entropy, in common with other soluble carbohydrate/protein interactions (Charnock *et al.*, 2000; Charnock *et al.*, 2002b; Boraston *et al.*, 2003; Henshaw *et al.*, 2004).

The ITC data shows that *Ct*CBM6 binds preferentially to xylan and xylooligosaccharides of at least two xylose moieties, exhibiting maximum affinity for the pentasaccharide. *Ct*CBM6 is thus primarily a xylan-binding module, and by analogy it is likely that the other members of family 6 CBMs, which are components of xylanases, also recognise xylan. These modules, however, are also found in GHs that act on a diversity of plant polysaccharides other than xylan. Thus, it is likely that, similar to the CBM families 2, 4 and 22, different members of CBM6 will recognise distinct polysaccharides. This view is supported by a study that showed that the N-terminal CBM6 in xylanase 11A from *C. stercorarium* binds specifically to xylan, whereas the other two CBM6 modules in this protein bind to cello- and xylooligosaccharides with similar affinity (Boraston *et al.*, 2003). Furthermore, the C-terminal CBM6 (*Cm*CBM6-2) in *C. mixtus* endoglucanase 5A binds to a variety of β-glucans but displays no significant affinity for xylan, although it binds weakly to xylohexaose (Henshaw *et al.*, 2004)

The stoichiometry for the binding of *Ct*CBM6 to xylopentaose is close to one indicating that the protein module contains one distinct ligand-binding site that can accommodate these oligosaccharides. This is similar to the stoichiometry of ligand binding of other CBMs, such as CBM4 (Tomme *et al.*, 1996), CBM9 (Boraston *et al.*, 2001a; 2001b), and CBM22 (Charnock *et al.*, 2000), which also contain single carbohydrate-binding sites, but is in contrast to the starch-binding modules of *Aspergillus* glucoamylase, which accommodates two amylose chains (Sorimachi *et al.*, 1997). Similarly, *Cm*CBM6-2 also contains two discrete ligand-binding sites (Henshaw *et al.*, 2004)

Enthalpic forces have been shown to promote ligand binding by *Cf*Xyn11A CBM2b-1 and *Cf*Cel9B CBM4 (Tomme *et al.*, 1996; Simpson *et al.*, 1999). The

thermodynamics of *Ct*CBM6-ligand interactions are similar to other proteins that bind soluble carbohydrates (Tomme *et al.*, 1996; Boraston *et al.*, 2001a; 2001b; Charnock *et al.*, 2000), whereas favourable entropic forces dominate the binding of CBM2a proteins to crystalline cellulose (Creagh *et al.*, 1996). When interacting with *Ct*CBM6, soluble xylan and xylooligosaccharides appear to encounter conformational restriction with the resultant loss in entropy, which is bigger than the positive entropy contribution from the release of water from both the protein and the carbohydrate. Therefore, the main forces that drive the interaction are enthalpic. The positive entropy derived from CBM2a-crystalline cellulose interactions likely reflects the release of ordered water molecules on the surface of the ligand and protein (Creagh *et al.*, 1996). The loss of conformational entropy normally associated with the binding of proteins to soluble carbohydrates does not occur as crystalline cellulose is not highly mobile.

4.7 Role of CBM6 in *Ct*Xyn11A

CBMs have been demonstrated to enhance the degradation rate by cellulases and xylanases against insoluble polysaccharides but not soluble substrates (Din *et al.*, 1991; 1994; Bolam *et al.*, 1998; Kuno *et al.*, 1998; Sun *et al.*, 1998; Fernandes *et al.*, 1999; Gill *et al.*, 1999). *Pf*Xyn10A CBM2a increased the activity of the xylanase against a complex of insoluble xylan and cellulose (Gill *et al.*, 1999). It was shown that *Cf*Cel6A CBM2a enhanced catalytic hydrolysis not only by promoting substrate attachment, but also by disrupting the structure of cellulose, making the substrate more accessible to enzyme attack (Din *et al.*, 1991; 1994). *Ct*CBM6 characterised in this chapter also displayed synergistic action with the CM against insoluble xylan (Fernandes *et al.*, 1999).

It is currently unclear what the exact role *Ct*CBM6 plays. It is possible that this module may increase the proximity of the substrate and the protein so as to influence the degradation of the natural polysaccharide.

It is unclear why *Ct*CBM6 exhibits this binding pattern, but it is arguable that the CBM may influence the degradation of xylans and arabinoxylans in the PCW by full length *Ct*Xyn11A. *Ct*CBM6 binds to the xylan backbone and can therefore increase the effective concentration of the enzyme on the side chains of xylan and

make the substrate more accessible to attack by the enzyme. This hypothesis is supported by the observation that *Pseudomonas* CBMs that have an affinity for cellulose mediate their effects on catalysis by increasing enzyme-substrate proximity (Bolam *et al.*, 1998; Gill *et al.*, 1999).

Interestingly, the interaction between *Ct*CBM6 and xylan is relatively unaffected by the degree of arabinose substitutions on the polysaccharide backbone. The arabinose component of oat spelt xylan, birchwood xylan, wheat arabinoxylan and rye arabinoxylan are 6 %, 10 %, 50 % and 70 %, respectively (Li *et al.*, 2000), and the protein showed similar affinity for varying degrees of substitution in different xylans.

It is interesting to note that *Ct*CBM6, as well as binding to unsubstituted xylooligosaccharides, is also able to interact with wheat and rye xylans, both of which are highly substituted at the C2 and C3 positions with arabinose residues (Joseleau *et al.*, 1992). Structural studies by Atkins (1992) have shown that the basic 3-fold screw axis of xylan is not significantly affected by the nature and extent of sugar substitution at the 2 and 3 positions on the xylose ring. This is highly significant in relation to *Ct*CBM6 in the full length enzyme, as in plant cell walls the majority of xylans are substituted, although the nature and extent of this decoration varies, depending on the origin and differentiation state of the plant material (Brett and Waldron, 1996). The observation that *Ct*CBM6 can accommodate both highly and poorly substituted xylans indicates that this module will also be able to bind to a wide range of different plant cell walls, and this flexibility in ligand specificity provides a possible explanation for the evolution of family 6 CBMs.

4.8 Ligand-binding in other CBM6s

It is likely that, similar to the CBM families 2 and 4, different members of CBM6 recognise distinct polysaccharides. This view is supported by a study by Boraston *et al.* (2001a; 2001b, 2003) that showed that the N-terminal CBM6 in xylanase 11A from *C. stercorarium* binds specifically to xylan, whereas the other two CBM6 modules in this protein bind to cello- and xylooligosaccharides with similar affinity.

The 3D structures of almost all Type B CBMs determined to date conform to a classic lectin-like β-jelly roll in which a single ligand-binding site comprises a shallow cleft on the concave surface of the protein. In CBM4 and CBM2, variation in ligand recognition between different members of these families is reflected in the structure of this conserved binding site (Boraston *et al.*, 2002a). While two xylan-binding CBM6s from *C. thermocellum* Xyn11A (*Ct*CBM6; this study) and *C. stercorarium* putative xylanase (*Cs*CBM6-3) also display a β-jelly roll fold with a potential ligand-binding site (Cleft B) on the concave surface, a second possible binding cleft (Cleft A) is located on one edge of the protein between the loops that connect the inner and outer β-sheets (Boraston *et al.*, 2003). However, *Cs*CBM6-3 studies have shown, as with *Ct*CBM6, that only cleft A binds ligands, while Cleft B is unable to interact with carbohydrates, as it appears to be partially occluded by a proline-containing loop, analogous to *Ct*CBM6 (Boraston *et al.*, 2003; Czjzek *et al.*, 2001).

An observation of this study is that proteins that share a common ancestor (as evident from high structural similarity) have ligand-binding sites in different locations on the protein scaffold. Thus, data presented in this chapter show that Cleft A is the ligand-binding site for *Ct*CBM6, whereas cleft B binds to the target oligosaccharides in CBM4 and CBM22 (Brun *et al.*, 2000; Xie *et al.*, 2001b), which share a common evolutionary origin with CBM6. This type of phenomenon is seen in CBM10 from *Pseudomonas* xylanase 10A, which binds to crystalline cellulose and has an oligonucleotide-binding fold, but the ligand-binding site in the CBM is in a different location to oligonucleotide-binding proteins that have the same fold (Raghothama *et al.* 2000).

Another interesting feature of this study is that the variation in the location of the functional binding-site in the different members of CBM6 may contribute to the diversity of ligand specificity seen with this family, which contains proteins that bind to cellulose and xylan. This view is supported by the lack of conservation, in family 6 members, of residues that are known to play a key role in ligand binding

in *Ct*CBM6 (Figure 3.3). This suggests that the CBM6 scaffold displays a remarkable structural and functional plasticity.

As with *Cs*CBM6-3, *Ct*CBM6 contains a surface loop that partially occludes Cleft B and thus prevents access to the ligand. It is possible that for other members of CBM6, the loop is in a different conformation, making Cleft B accessible to xylan and xylooligosaccharides. Support for this view comes from a recent study by Henshaw *et al.* (2004) on *Cm*CBM6-2. The structure of *Cm*CBM6-2 shows that the corresponding "clostridial proline-loop" is much smaller, compared to those in *Ct*CBM6 and *Cs*CBM6-3, and lacks the ridigity conferred by proline, thus allowing Cleft B in the *Cellvibrio* protein to bind to saccharide chains (Henshaw *et al.*, 2004; Pires *et al.*, 2004). The data showed that the protein has two functional binding clefts and can accommodate β-1,4- and β-1,3-linked glucans in Cleft A and B and that the two binding sites act in synergy to bind insoluble cellulose. Only Cleft B, however, interacts with β-1,3-1,4-mixed linked glucans such as lichenan and barley β-glucan, supporting the hypothesis that the two potential binding clefts contribute to the variation in ligand specificity observed within the CBM6 family (Henshaw *et al.*, 2004).

4.9 Summary

To summarise, CBMs with an affinity for xylan have been shown to be present in xylan-degrading enzymes including xylanases of GH families 10 and 11, and arabinofuranosidases from several organisms. *Clostridia* enzymes contain functional CBMs that exhibit distinct ligand specificity. Although the interactions between CBMs and their ligands are different, they share some common features such as hydrophobic stacking interactions between surface aromatic residues and sugar rings, and hydrogen bonds between the polar amino acids and the hydroxyl groups of the sugar polymer. The structures of CBMs, which interact with xylan, solved to date are similar, comprising clefts rather than a flat binding surface found in CBMs that interact with crystalline cellulose (Xu *et al.*, 1995; Raghothama *et al.*, 2000). It can be concluded that the interaction between *Ct*CBM6 and xylan shares the common mechanism of carbohydrate/protein association; that is hydrophobic forces and hydrogen-bonds.

CBMs with an affinity for xylan exist in xylan-degrading enzymes from *C. thermocellum*. *Ct*CBM6 is a functional CBM that binds with equal affinity to highly decorated or poorly substituted xylans. At least two β-1,4-linked xylopyranosides are required for *Ct*CBM6 to bind xylose polymers, and it is likely that the surface of *Ct*CBM6 accommodates five xylose residues. The interaction between *Ct*CBM6 and its ligands is relatively strong, with an association constant of 3.7×10^3 M^{-1} for xylobiose and 189.0×10^3 M^{-1} for xylohexaose. The module is thus a xylan-binding module.

*Ct*CBM6 binds to its ligands by a combination of hydrophobic stacking interactions and hydrogen bonding forces. Tyr-34 and Trp-92 stack to the two faces of a single xylose moiety. In contrast, Asn-120 is not only structurally important but also is likely to make a direct contribution to ligand binding through hydrogen bonding interactions, via its hydroxyl group on its side chain to sugar hydroxyl groups. Indeed, a recent paper by Pires *et al.* (2004) has shown that Asn-120 makes strong hydrogen bonds with the O3 and O5 of adjacent xylose residues in the xylopentaose ligand, while Trp-92 and Tyr-34 stack against the two faces of the same xylose unit. Pires *et al.* (2004) have recently solved the structure of *Ct*CBM6 to 1.6 Å with xylopentaose in complex (Figure 4.1), which occupies the complete binding site. This provides an insight into the mechanism by which Cleft A in family 6 CBMs can distinguish between *gluco-* and *xylo-* configured ligands. Xylopentaose has been shown to bind in an almost 3-fold helical orientation characterised by the internal hydrogen bond from O3 and O5 of adjacent xylose molecules (Pires *et al.*, 2004). Such a conformation has previously been proposed for xylohexaose bound to CBM2b-1 from *C. fimi* Xyn11A (Simpson *et al.*, 2000), and observed when the ligand is bound to CBM15 from *C. japonicus* Xyn10C (Szabó *et al.*, 2001).

The fact that *Ct*CBM6 bound to different forms of soluble xylans implies that the binding site of this module can accommodate linear xylan and xylan backbones that contain extensive side chains. AGE and ITC (Tables 3.3 and 3.4) revealed

that *Ct*CBM6 displayed similar affinity for xylans derived from oat spelt, birchwood, wheat and rye, which is similar to the binding specificity of

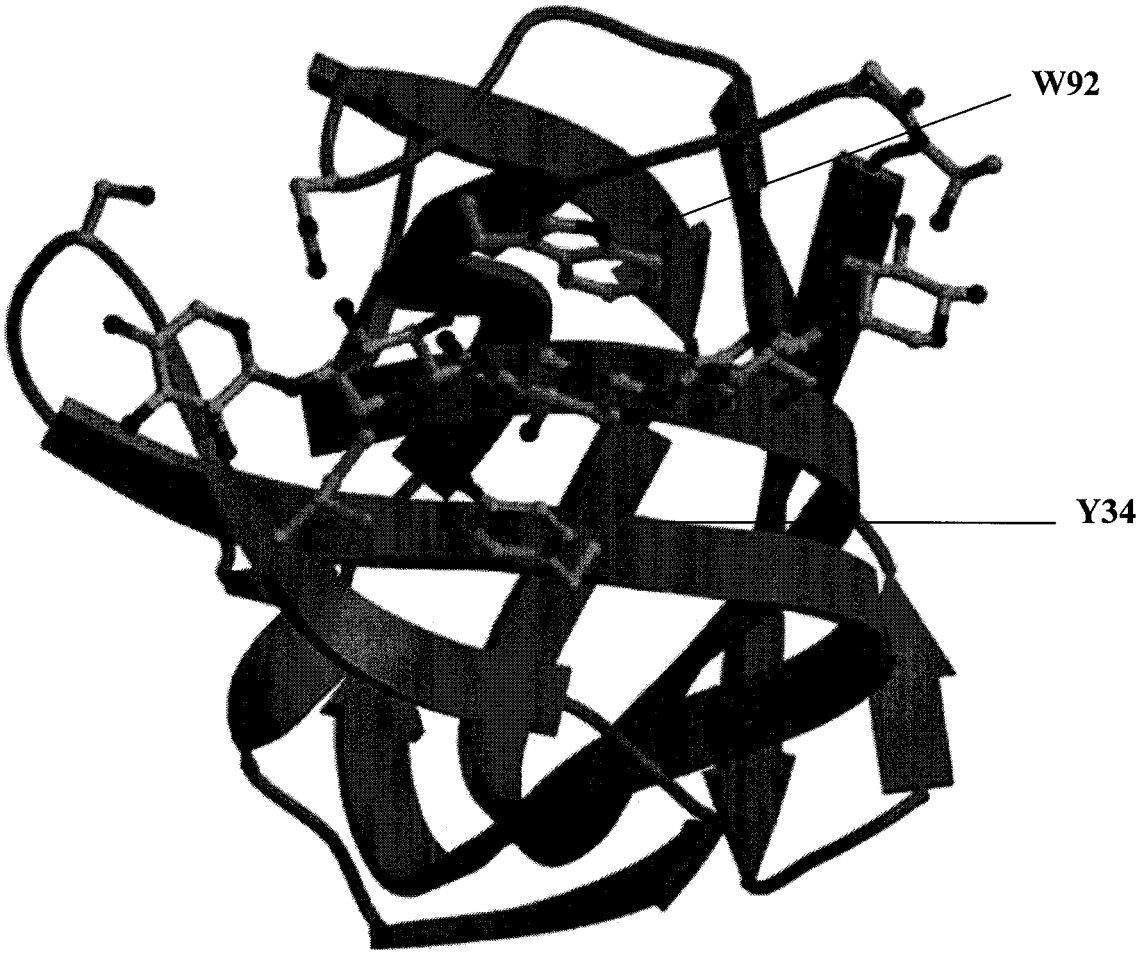


Figure 4.1: Xylopentaose bound to cleft A of *Ct*CBM6, represented as sticks.
The residues interacting with the ligand are highlighted. Taken from Pires *et al.*, 2004.

*Cf*Xyn11A CBM2b-1 (Simpson *et al.*, 1999) and Xyn10C CBM15 (Szabó *et al.*, 2001). Both of these CBMs interacted with highly substituted arabinoxylans, and oat spelt and birchwood xylan, which contain few arabinose side chains. The *Ct*CBM6 module exhibited weak affinity for cellohexaose, cellopentaose and HEC, CMC and methyl cellulose (Table 3.2). Data from both AGE and ITC showed that *Ct*CBM6 interacted much more strongly with xylan than with cellulose (Figures 3.7, 3.9-3.10, 3.18-3.19; Table 3.2, 3.4). These data suggest that *Ct*Xyn11A makes contact primarily with xylan chains, but its binding surface has room for different xylan side chains.

Structural studies showed that xylan-binding in *Ct*CBM6 contains two distinct clefts that could comprise the ligand-binding site. Mutagenesis, structural and NMR data showed that only one of these sites, designated Cleft A, could accommodate xylooligosaccharides or xylan, while the other cleft, which was unable to interact with the target ligands, was defined as Cleft B.

4.10 Future work

1. Further mutagenesis studies on *Ct*CBM6, to elucidate the role other key amino acids in Cleft A that may be involved in ligand binding.
2. Further crystallisation studies of mutant forms of *Ct*CBM6 co-crystallised with oligosaccharides.
3. Further biochemical characterisation of mutants using ITC.
4. The role of *Ct*CBM6 in the hydrolysis of substrates by full length *Ct*Xyn11A.
5. Remove the proline loop to ascertain whether Cleft B becomes functional.

Biochemical characterisation and crystallisation studies of three hyaluronidases from GH69.

5 Results

5.1 Introduction

Hyaluronan (hyaluronic acid; HA) is a high molecular weight linear polysaccharide composed of a repeating unit of β-1,3-D-glucuronic acid and β-1,4-*N*-acetyl-D-glucosamine (Table 1.1; Kreil, 1995; Laurent *et al.*, 1996; Fraser *et al.*, 1997). HA plays a central role in the formation and stability of the extracellular matrix (ECM), for example, in cartilage, skin and brain of vertebrates, via its interactions with HA-binding proteins and receptors (Day, 1999). Although the majority of HA is found in all tissues and body fluids of vertebrates, some bacteria also produce high amounts of HA, for example, *Streptococcus pyogenes*, *Streptococcus zooepidemicus*, *Streptomyces hyalurolyticus* (Fraser *et al.*, 1997; Menzel and Farr, 1998).

HA can form highly viscous solutions and thereby influence the properties of the ECM (Kreil, 1995). It has been implicated in many biological processes, including fertilisation, embryonic development, cell migration and differentiation, wound healing, inflammation, and growth and metastasis of tumour cells (Laurent and Fraser, 1992; Kreil, 1995; Laurent *et al.*, 1996). In general, it is present whenever rapid repair and regeneration occurs. Being a negatively charged molecule, the HA polymer takes on considerable amounts of water. In this way, it creates space between the cell, decreases cell-cell contacts, loosens tissues, by and large, pre-requisites for cellular migration and altered cellular communication (Laurent *et al.*, 1996; Fraser *et al.*, 1997).

5.1.1 HA-degrading enzymes

The term hyaluronidase was introduced to denote enzymes that degrade HA (Menzel and Farr, 1998). However, as mentioned in Section 1.3.1.1.7.1, from the literature, the term hyaluronidase is used to denote both glycoside hydrolases (GHs) and polysaccharide lyases (PLs) and therefore the term HA-degrading enzymes would be more appropriate. For the purposes of this thesis, the term

hyaluronidase will be used to denote the GHs and hyaluronate lyases will denote the PLs. Hyaluronidases are found in prokaryotes and eukaryotes, belonging to families GH56, GH69, GH84 and PL8 (CAZy).

In mammals, HA-degrading enzymes have been found in higher quantities in testis, liver lysosomes and serum. They are involved in controlling HA levels and are thus implicated in various diseases related to defects of HA metabolism, such as arthritis or tumour metastasis (Menzel and Farr, 1998). At least six genes potentially coding for hyaluronidases are present in the human genome (Csóka *et al.*, 2001).

Hyaluronidases are found in bacteria, such as *Streptococcus pyogenes*, and are thought to be required for the degradation of the HA capsule that surrounds the bacterium and also for aiding the spread of the organism through host connective tissue by hydrolysis of the HA component. HA-degrading enzymes are also found in the venom of insects, such as bees, wasps and hornets, as well as in the venom of some snakes and lizards (Meyer, 1971; Frost *et al.*, 1996; Csóka *et al.*, 1997).

5.1.2 *Streptococcus pyogenes*

S. pyogenes (Group A streptococcus) is an important Gram-positive, extracellular, bacterial pathogen (Cunningham, 2000). As pathogens, group A streptococci have developed complex virulence mechanisms to avoid host defences (Cunningham, 2000; Bisno *et al.*, 2003). *S. pyogenes* is distinguished from other streptococci by the serological grouping methods of Rebecca Lancefield, using immunological differences in their cell wall polysaccharides (Groups A, B, C, F and G) or lipoteichoic acid (Group D; Lancefield, 1928; Cunningham, 2000). The group A carbohydrate antigen is composed of *N*-acetyl-β-D-glucosamine linked to a polymeric rhamnose backbone (Cunningham, 2000; Figure 5.1). Numerous 'Lancefield groups' are now recognised, named by the letters A to J and K to T, and though many of these groups include several species, *S. pyogenes* is the only species of the group A streptococci (Bisno, 1979). Lancefield also showed that further subdividing of the group A streptococci into specific serological types was possible based on the presence of type-specific M (protein) antigens (Rotta *et al.*, 1971). Streptococcal M protein, which extends from the cell membrane of group

A streptococci (Figure 5.1), has been used to divide *S. pyogenes* into serotypes. Lancefield called these type-specific antigens M proteins, after the matt appearance of colonies in which they were first found, and showed that they are proteins by their sensitivity to proteases such as trypsin. Over 80 different serotypes of M proteins, usually called M1, M2, M3.....etc.

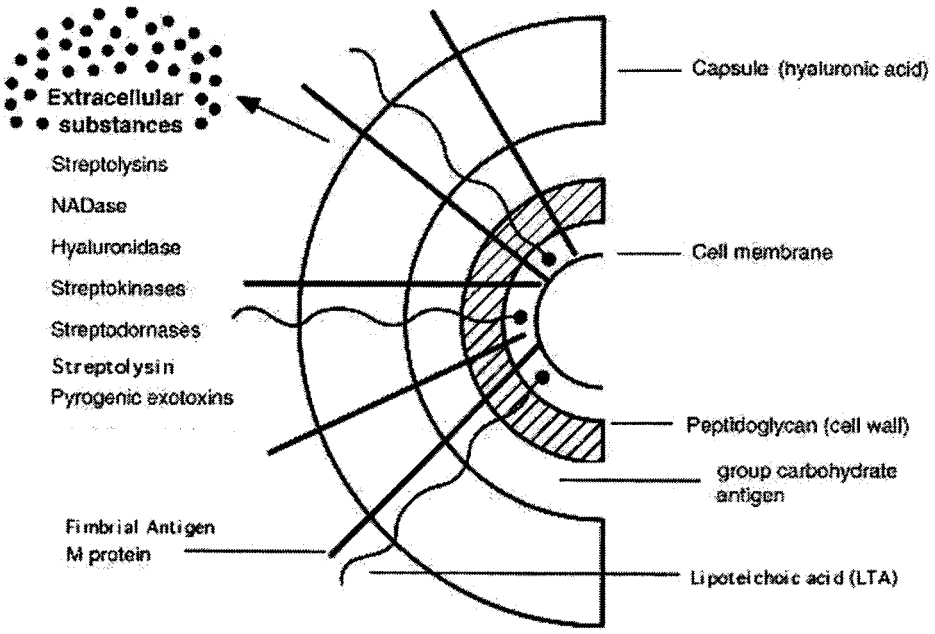


Figure 5.1: Cell surface structure of *Streptococcus pyogenes* and secreted products involved in virulence.
Taken from http://www.bact.wisc.edu/Bact330/lecturespyo.

have been described. Until recently, it was assumed that the M typing patterns are generally consistent and that they identify specific strains (Johnson and Kaplan, 1993). However, in recent years it has been shown that M antigen genes have often been transferred horizontally between different strains so that a single M type can include very different *S. pyogenes* strains (Whatmore and Kehoe, 1994; Whatmore *et al.*, 1994). This questions the reliability of identifying strains by the above typing systems and the reliability of many epidemiological studies that associated particular M types of *S. pyogenes* with particular diseases (Kehoe *et al.*, 1996).

5.1.3 Diseases caused by group A streptococci (GAS)

Group A streptococci are extracellular bacterial pathogens that produce a wide variety of diseases, primarily localised infections of the mucous membranes, tonsils or the skin, but occasionally invading deeper sites, causing pharyngitis, impetigo/pyoderma, cellulitis, necrotising fasciitis, toxic streptococcal shock syndrome, scarlet fever, septicaemia, pneumonia and meningitis (Cunningham, 2000; Bisno *et al.*, 2003). Group A streptococci are the most common bacterial cause of pharyngitis and primarily affect school-age children (Cunningham, 2000). Pharyngitis ranges from sub-clinical or mild throat infection to more severe infections producing symptoms including malaise, fever, nausea and inflammation of the pharyngeal mucosa (Gray, 1991). *S. pyogenes* can spread to adjacent tissues to produce tonsillitis, otitis media and sinusitis, and it sometimes invades to cause septicaemia. Infections of the skin include impetigo, a localised infection that begins where the skin has been broken or damaged. Further spread of *S. pyogenes* may produce erysipelas, a diffuse inflammation of the skin and subcutaneous tissue. As with pharyngeal infections, *S. pyogenes* infecting the skin can also invade to cause septicaemia (Stevens, 1992). In some patients, *S. pyogenes* infections can lead to nonsuppurative sequelae, namely acute rheumatic fever and post-streptococcal acute glomerulonephritis. Acute rheumatic fever occurs only after pharyngeal infections, which might by asymptomatic, whereas post-streptococcal acute glomerulonephritis may occur after pyogenic infections of either the throat or skin. These sequelae are the result, at least in part, of the immune response to group A streptococci. Post-streptococcal acute glomerulonephritis presents the distinctive symptoms of uraemia, oedema, and raised blood pressure, and hence is quite easily diagnosed (Gray, 1991). Symptoms of acute rheumatic fever include fever, polyarthritis, and carditis (Homer and Shulman, 1991). Strains of *S. pyogenes* that produce streptococcal pyrogenic exotoxins, also called erythrogenic toxins, can cause toxic-shock syndromes and scarlet fever. In scarlet fever, a hypersensitivity reaction to these toxins gives rise to a red rash. Toxic shock is caused by the superantigenic activity of the toxins and the symptoms include fever, hypertension, and a scarlatinal form rash (Gray, 1991). There has been an increase in sporadic cases of severe GAS

toxic shock syndrome (Stevens, 1992). Many of these cases were associated with severe invasive tissue infections such as necrotizing fasciitis, while others followed apparently mild pharyngitis (Stevens, 1994). Since the incidence of GAS pharyngitis has not shown a concurrent increase in prevalence, it is thought that raised virulence and spread of specific strains of streptococci may be responsible for the upsurge of invasive GAS disease. GAS infections therefore remain a worldwide problem and continuing research is needed to keep pace with this evolving pathogen. A wide variety of cell-surface and extracellular factors are known or suspected to contribute to the virulence of *S. pyogenes* (Bisno *et al.*, 2003).

5.1.4 Extracellular virulence factors

5.1.4.1 Pyrogenic exotoxins

The first identification of a streptococcal pyrogenic toxin (Spe) was Dick and Dick's discovery of the erythrogenic toxin (1924). SpeA and SpeC are encoded by bacteriophage and belong to the family of exotoxins that includes the staphylococcal enterotoxins and toxic shock syndrome toxin, TSST-1 (Betley *et al.*, 1992). As these toxins are all pyrogenic, they have a close association with scarlet fever and toxic shock syndrome (Bisno *et al.*, 2003). Other Spes have been identified from genome sequence information. SpeB is a cysteine protease and expressed from a chromosomal gene in almost all group A streptococci strains (Betley *et al.*, 1992; Cunningham, 2000; Bisno *et al.*, 2003). SpeA and SpeC have superantigenic activity, like the staphylococcal enterotoxins and TSST-1, and they share significant primary sequence homologies with these staphylococcal toxins (Betley *et al.*, 1992). However, contrary to earlier reports, SpeB does not appear to be a superantigenic toxin, and it shares no significant sequence homology with either the staphylococcal superantigenic toxins or with SpeA and SpeC. It is thought that SpeB's contribution to disease results solely from the protease activity (Betley *et al.*, 1992) by promoting tissue damage by degrading fibronectin and vitronectin in the host extracellular matrix (Kapur *et al.*, 1993).

5.1.4.2 Cytolytic toxins

Most group A streptococci express two different cytolytic toxins called streptolysin O (SLO) and streptolysin S (SLS). SLO, which derives its name from

its oxygen lability, is a member of the thiol-activated family of toxins (Alouf, 1980). SLO binds to target eukaryotic membranes, such as erythrocytes, polymorphonuclear leukocytes, platelets and lysosomes, and forms pores in the membrane through which large cytoplasmic molecules escape and hence cell lysis occurs (Alouf, 1980; Bisno *et al.*, 2003). SLO is produced *in vivo* during *S. pyogenes* infections, as indicated by an increase in the levels of anti-SLO antibodies, suggesting it plays a role in virulence (Alouf, 1980).

SLS is a non-antigenic haemolysin and it is very unstable when it is separated from large "carrier" molecules such as serum proteins. It exists in intracellular, cell-surface bound and extracellular forms and is one of the most potent cytotoxins known (Dale *et al.*, 2002; Bisno *et al.*, 2003). SLS is cytolytic for a great variety of eukaryotic cells, with the capacity to damage the membranes of polymorphonuclear leukocytes, platelets and subcellular organelles (Alouf and Loridan, 1988). SLS causes ions to leak from the cell, resulting in cell lysis by an osmotic mechanism (Bernheimer, 1972). Unlike SLO, SLS is not inactivated by oxygen, but it is quite thermolabile (Bisno *et al.*, 2003).

5.1.4.3 Hyaluronidase

Hyaluronidase may aid in the spread of group A streptococci in the host by breaking down the hyaluronic acid found in connective tissues. However, both of the genes cloned and sequenced so far, *hylP* (Hynes and Ferretti, 1989) and *hylP2* (Hynes *et al.*, 1995) are encoded by streptococcal bacteriophages, which may have evolved these genes to facilitate phage infection of strains with capsules of hyaluronic acid, and lack a recognisable signal sequence. Extracellular hyaluronidase may therefore be either encoded by a chromosomal gene, unlike the sequenced phage genes, which includes a signal sequence, or released only upon lysis of phage-infected streptococci.

5.1.4.4 Streptokinase

Streptokinase is secreted by several species of haemolytic streptococci and is of therapeutic interest as a plasminogen activator (Martin, 1982). It is not itself a protease, but it promotes the dissolution of blood clots by catalysing the conversion of plasminogen to plasmin that degrade fibrin (Martin, 1982). Its

contribution to streptococcal virulence may be the inhibition of the deposition of fibrin barriers around infected lesions.

5.1.5 Cell surface virulence factors

Lancefield identified several of the cell surface proteins of *S. pyogenes* and studied their antigenicity. The most important of these is M protein. Other group A streptococci cell surface proteins that are thought to contribute to virulence include lipoteichoic acids (LTA), fibronectin-binding protein (Protein F) and the capsule.

5.1.5.1 M protein

The major human host defence against invasive group A streptococcal infection is that of phagocytosis and killing by polymorphonuclear leukocytes (Bisno *et al.*, 2003). Thus, a critical group A streptococcal virulence factor is an antiphagocytic surface constituent known as M protein, which was originally identified by Lancefield (1928).

In a review by Lancefield (1962), the M antigen is described as a major virulence factor that inhibits the phagocytosis of M^+ streptococci by polymorphonucleocytes (PMNLs). Lancefield found that immunity to GAS is largely due to anti-M protein antibodies, which neutralise the antiphagocytic properties of M protein by opsonising the streptococcal cells. However, these antibodies often cannot protect against subsequent streptococcal infections since there are over 80 different serotypes of M antigen and protective antibodies are largely M type-specific with very limited cross-protection (Lancefield, 1959).

Lancefield's work showed that M protein could be proteolytically removed from streptococcal cells, which subsequently retained both viability and the ability to resynthesise M protein. This suggested that M protein was situated on the cell wall (Lancefield, 1962). M protein is anchored in the cell membrane, traversing the cell wall, and appears as fibrils on the cell surface (Cunningham, 2000; Figure 5.1).

The adherence of group A streptococci to epithelial surfaces allows them to colonise host surfaces, for example, in the buccal and pharyngeal cavity, which is continually bathed with saliva or imbibed fluids. In view of the requirement of M proteins for virulence and their prominent cell surface location, they were considered likely to have a role in adherence. Several studies into the potential role of M protein and other streptococcal cell surface moieties as adhesins, particularly LTA and fibronectin-binding proteins, produced apparently conflicting results. This may simply reflect the fact that streptococcal adhesin is complex, relying on several adhesins whose roles vary with the phenotype of the infecting strain, the type of host cell involved, and other environmental factors. Early studies implicated M protein as a mediator of group A streptococci adherence, but subsequent work has shown that M protein does not promote adhesion to epithelial cells (Caparon *et al.*, 1991). Group A streptococci surface proteins that bind fibronectin, such as LTA, which is anchored to proteins on the bacterial surface, and fibronectin-binding proteins (Protein F), have been shown to be important in adherence to both throat and skin (Bisno *et al.*, 2003; Figure 5.1).

5.1.5.2 Capsule

A number of possible functions have been suggested for polysaccharide capsules. The formation by capsules of a hydrated gel around the surface of the bacteria may protect the bacteria from the harmful effects of dessication (Roberson and Firestone, 1992). This may be particularly relevant in aiding the transmission of encapsulated pathogens from one host to the next (Roberts, 1996). Capsular polysaccharides may promote the adherence of bacteria to both surfaces and to each other (Roberts, 1996). During invasive bacterial infections, interactions between the capsular polysaccharide and the host's immune system can decide the outcome of the infection (Roberts *et al.*, 1989). In the absence of specific antibodies, the presence of a capsule is thought to confer resistence to non-specific host defence mechanisms (Roberts, 1996). Although most capsular polysaccharides can elicit an immune response, a small set of capsular polysaccharides are poorly immunogenic, such as group A streptococcal capsular hyaluronate (Roberts, 1996; Bisno *et al.*, 2003). As a consequence of structural similarities between these capsular polysaccharides and polysaccharides encountered on host tissue, the capsules are poorly immunogenic, and infected

individuals mount a poor antibody response to such capsules (Roberts *et al.*, 1989; Bisno *et al.*, 2003). Therefore, the expression of these capsules confers some measure of resistance to the host's specific immune response (Roberts, 1996).

The group A streptococcal capsule is composed of hyaluronic acid (Figure 5.1), which is composed of alternating residues of *N*-acetylglucosamine and glucuronic acid. The capsule, like M protein, is considered as a virulence factor, conferring antiphagocytic properties upon the streptococcal cell (Foley and Wood, 1959; Moses *et al.*, 1997). In addition, the capsule may be an important adherence factor in the pharynx, since it binds CD44 on epithelial cells (Schrager *et al.*, 1998; Knudson *et al.*, 2002).

The production of hyaluronic acid is controlled by an operon composed of three different genes: *hasA*, *hasB* and *hasC* (Crater and vin de Rijn, 1995). Streptococcal strains vary greatly in their degree of encapsulation, and those with the most prolific production have a mucoid appearance (Whitnack *et al.*, 1981; Bisno *et al.*, 2003). Group A streptococcal strains rich in M protein and capsule are extremely virulent to humans, and mucoid strains have long been seen to cause deeply invasive infections (Pilot, 1944; Stollerman, 1996; Bisno *et al.*, 2003). Group A streptococcal capsular hyaluronate is chemically quite similar to that found in human connective tissue. For this reason it is a poor immunogen and antibodies to group A streptococcal hyaluronate have been quite difficult to demonstrate in people (Bisno *et al.*, 2003).

5.1.5.3 Genome sequence

Ferretti *et al.* (2001) completed the entire sequence of the group A streptococcal genome from *S. pyogenes* SF370 strain M1. The bacterial chromosome of strain M1 contains genes encoding for the hyaluronic acid capsule (*has*) operon coding for the hyaluronidase capsule. The genome sequence confirmed the observation that bacteriophages are greatly important in genetic variation, and that introduction of genes encoding toxins and other virulence-associated proteins is often via phage transduction (Wagner and Waldor, 2002). Phage DNA accounts for 7.1 % of the M1 genome, and codes for three hyaluronidases (Ferretti *et al.*, 2001).

Several putative genes encoding proteins with internal repeats of the motif sequence Gly-X-Y were identified in the *S. pyogenes* SF370 genome (Ferretti *et al.*, 2001). These amino acid triplet repeats resemble the characteristic repeating sequences found in collagen. Of interest is the presence of this collagen repeat in ORFs HylP2 and HylP3. In these enzymes, 10 repeats of the collagenous Gly-X-Y repeat are found, with four proline residues in the X-Y position (Figure 5.2). The 10 triplet collagen repeat is the minimum length required to provide stability for a triple helical structure (Berisio *et al.*, 2003). Sequence analysis also showed that this collagen repeat is also found in *S. pyogenes* bacteriophage H4489A (Figure 5.2; Hynes and Ferretti, 1989), but as yet the function of these sequences is unknown.

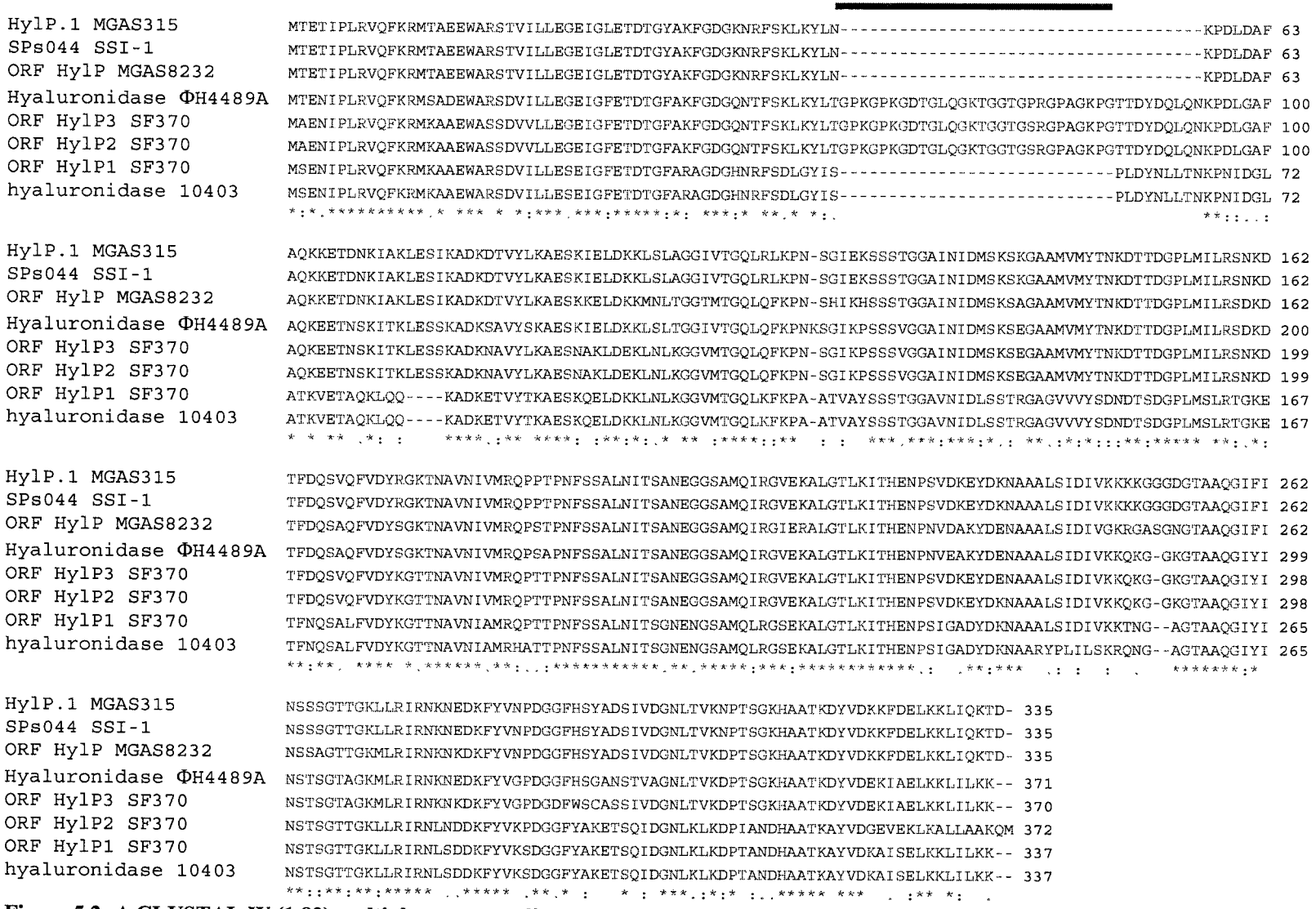


Figure 5.2: A CLUSTAL W (1.82) multiple sequence alignment of a representative sample of GH69 open reading frames (ORFs).
HylP.1 from *S. pyogenes* MGAS315; SPs044 from *S. pyogenes* SSI-1; ORF HylP from *S. pyogenes* MGAS8232; hyaluronidase from *S. pyogenes* phage H4489A; HylP1, HylP2 and HylP3 from *S. pyogenes* SF370; and a hyaluronidase from *S. pyogenes* 10403. Several putative genes encoding proteins with internal repeats of the motif sequence Gly-X-Y were identified in the *S. pyogenes* SF370 genome (Ferretti *et al.*, 2001). These amino acid triplet repeats resemble the characteristic repeating sequences found in collagen. Sequence analysis showed ORFs HylP1, HylP2 and HylP3 from *S. pyogenes* SF370 and *S. pyogenes* phage H4489A to contain this collagen repeat (red line), for which the function of these sequences is as yet unknown.

This chapter will describe the biochemical properties of three structurally related enzymes, HylP1, HylP2 and HylP3, of *S. pyogenes* SF370. HylP1, HylP2 and HylP3 belong to the GH69, a family for which there is limited biochemical information, with only two enzymes described to date (CAZy). Hence, much of this results section is dedicated to the biochemical characterisation of HylP1, HylP2 and HylP3, answering questions such as pH optimum, temperature optimum, cofactor requirements, substrate preference, mode of action and absolute kinetic parameters (K_M and k_{cat}). Although the primary aim of the work undertaken in this chapter was to clone and investigate the biochemical parameters of HylP1, HylP2 and HylP3 GH69 from *S. pyogenes* SF370, data will also be shown regarding crystallographic studies that were carried out to solve the 3D structure of HylP1, which will allow the structure determination of HylP2 and HylP3 by molecular replacement.

The results show that ORFs HylP1, HylP2 and HylP3, encoding putative hyaluronidases, from *S. pyogenes* SF370 were successfully cloned and expressed in *Escherichia coli* and purified. ORFs HylP1, HylP2 and HylP3 were shown to be catalytically active on the substrate hyaluronate and have therefore been assigned as hyaluronidases. Characterisation of the N-terminally hexahistidine tagged HylP1 (38.4 kDa), HylP2 (42.0 kDa) and HylP3 (41.8 kDa) enzymes revealed an optimum pH of 6.5, 6.0 and 5.5 respectively. At 37 °C, HylP1, HylP2 and HylP3 displayed the highest amount of activity and overall, none of the enzymes displayed an absolute requirement for any of the divalent ions tested. K_M for HylP1, HylP2 and HylP3 has been determined as 0.90, 2.07 and 4.35 mg ml^{-1}, with a k_{cat} of 1390.90, 742.01 and 1253.04 s^{-1}, respectively, against sodium hyaluronate. Moreover, PAGE analysis showed each enzyme was endo-cleaving, generating a range of oligosaccharides.

All three enzymes have been crystallised and sufficient quality diffraction data obtained for native HylP1 and HylP3 (data collection and processing was performed by Dr Edward Taylor, YSBL) enabling the determination of crystal parameters: spacegroup H32 and unit cell dimensions of (*a*) 58.53 Å, (*b*) 58.53 Å and (*c*) 86.58 Å for native HylP1 and spacegroup P622 and unit cell dimensions

of (*a*) 49.48 Å, (*b*) 49.48 Å and (*c*) 241.0 Å for native HylP3. The 3D structure of HylP1 has been solved at a resolution of 1.8 Å and is composed of three monomeric strands, comprising α- and β-helical strands that are intertwined, giving rise to the triple-helical nature of the protein. The trimeric structure of HylP1 resembles a phage tail protein.

5.2 Protein expression and purification

5.2.1 Cloning, expression and purification of HylP1, HylP2 & HylP3 native recombinant protein

To investigate the biochemical properties of HylP1, HylP2 and HylP3 from glycoside hydrolase family 69 (GH69), the native recombinant protein was cloned and expressed. The segment of the *HylP1, HylP2,* and *HylP3* gene encoding the putative hyaluronidases from *S. pyogenes* SF370 were amplified by PCR from the genomic DNA of *S. pyogenes* SF370. Different primer pairs (Table 2.7) were used to generate two PCR products for each gene encoding HylP1, HylP2 and HylP3 (Stop and No stop; Figure 5.3 A; Section 2.2.2.1.2). The amplified DNA was then digested with the appropriate restriction endonucleases (Figure 5.3 B; Section 2.2.2.1.3) and cloned into similarly restricted pET-22b or pET-28a expression vectors (Section 2.2.2.1.3) to generate plasmids that encode the protein of interest with a flanking region on the fragment that codes for a His_6 tag at its C- or N-terminus or no His_6 tag. This was to increase the chance of obtaining soluble, active protein for HylP1, HylP2 and HylP3. These cloned products were then transformed into *Escherichia coli* super competent XL10-Gold cells, yielding ~ 50 colonies for each construct (Section 2.2.2.1.13.1). Plasmid DNA from these pET constructs (purified from one colony) was made (Section 2.2.2.1.6.2) and sent for sequencing using the standard sequencing primers T7 promoter and T7 terminator (Section 2.2.2.1.10) to confirm the correct gene had been amplified and no mutations were present. Native recombinant HylP1, HylP2 and HylP3 plasmid DNA was then chemically transformed into *E. coli* BL21 chemically competent cells (DE3; Section 2.2.2.1.13.1) and the recombinant *E. coli* strain was grown at 37 °C to mid-exponential phase. Expression of HylP1, HylP2 and HylP3 was induced by the addition of 1 mM IPTG (final concentration) and the cultures incubated at 30 °C overnight (Section 2.2.1.2.3). Cells were harvested by

centrifugation and cell free extracts (CFEs) were prepared as described in section 2.2.3.3. From analysis of the CFEs of HylP1, HylP2 and HylP3 containing a C- or N-terminal His_6 tag or no His_6 tag, it was observed that soluble protein was only obtained when an N-terminal His_6 tag was present for HylP1 (Data not shown). In contrast, soluble protein was obtained for all three contructs of HylP2 and HylP3 (Data not shown). However, it was decided that the N-terminally tagged recombinant proteins for HylP2 and HylP3 would be used for further studies.

The proteins were purified from cell free extracts by immobilised metal affinity chromatography (IMAC) under non-denaturing conditions, using a chelating sepharose column and an automated gradient elution technique (Section 2.2.3.4.2). The proteins were checked for purity by SDS-PAGE (Section 2.2.3.1; Figure 5.4), their concentrations determined by Bradford's assay (Section 2.2.3.7.2) and subjected to biochemical analysis (Section 2.2.5) and crystallisation studies (Section 2.2.3.8).

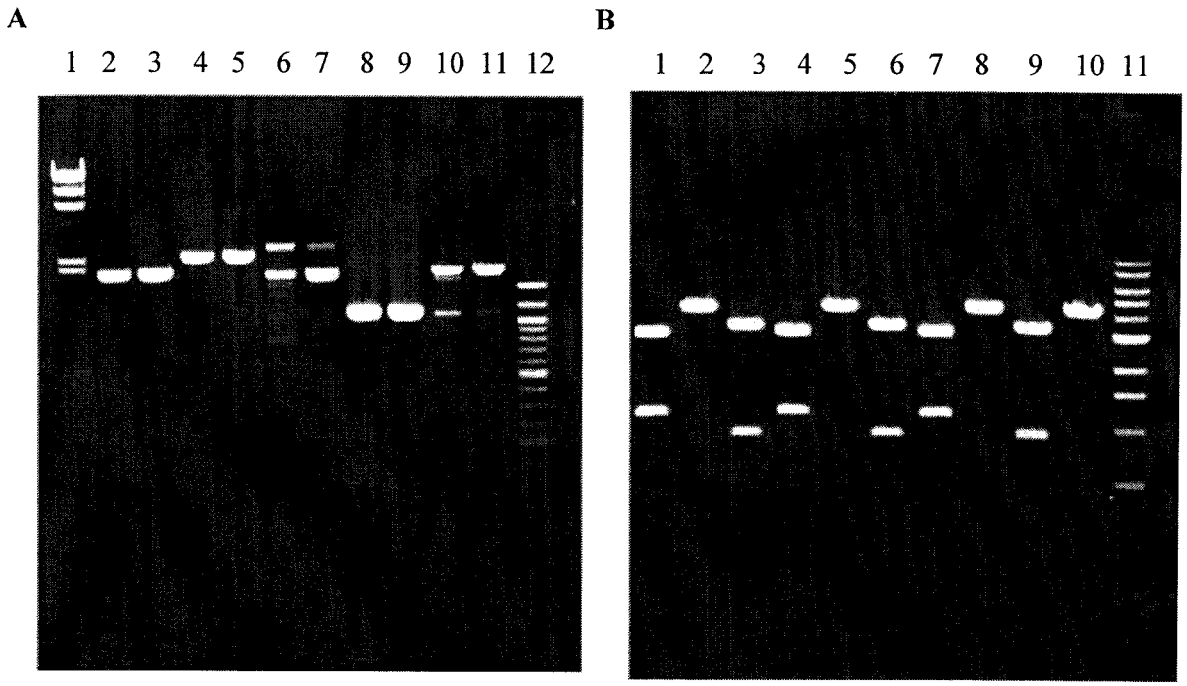


Figure 5.3: Visualisation of DNA bands following agarose gel electrophoresis. (A) DNA products generated during a PCR (Section 2.2.2.1.2). Lane 1: 1μg λDNA / *Hind* III ladder (23130, 9416, 6557, 4361, 2322, 2027, 564, 125 bp); lane 2: Spy0628 stop gene (PL12; 1908 bp) from *S. pyogenes* SF370; lane 3: Spy0628 no stop gene (PL12; 1908 bp) from *S. pyogenes* SF370; lane 4: HylA stop gene (PL8; 2418 bp) from *S. pyogenes* SF370; lane 5: HylA no stop gene (PL8; 2418 bp) from *S. pyogenes* SF370; lane 6: HylP1 stop gene (GH69; 1800 bp) from *S. pyogenes* SF370; lane 7: HylP1 no stop gene (GH69; 1800 bp) from *S. pyogenes* SF370; lane 8: HylP2 stop gene (GH69, 1119 bp) from *S. pyogenes* SF370; lane 9: HylP2 no stop gene (GH69; 1119 bp) from *S. pyogenes* SF370; lane 10: HylP3 stop gene (GH69; 1900 bp) from *S. pyogenes* SF370; lane 11: HylP3 no stop gene (GH69; 1900 bp) from *S. pyogenes* SF370; lane 12: 100 bp DNA ladder (1500, 1000, 900, 800, 700, 600, 500, 400, 300, 200, 100 bp). (B) DNA fragments generated during a restriction endonuclease digest (Section 2.2.2.1.3). Approximately 0.3 μg of each DNA was cut during a digest. Lane 1: HylP1 stop gene (GH69) restricted with *Nde* I; lane 2: HylP1 stop gene (GH69) restricted with *Xho* I; lane 3: HylP1 stop gene (GH69) restricted with *Sal* I; lane 4: HylP2 stop gene (GH69) restricted with *Nde* I; lane 5: HylP2 stop gene (GH69) restricted with *Xho* I; lane 6: HylP2 stop gene (GH69) restricted with *Hind* III; lane 7: HylP3 stop gene (GH69) restricted with *Nde* I; lane 8: HylP3 stop gene (GH69) restricted with *Xho* I; lane 9: HylP3 stop gene (GH69) restricted with *Eco*R V; lane 10: HylP1 no stop gene (GH69) restricted with *Nde* I; lane 11: 1 Kb DNA ladder (10, 8, 6, 5, 4, 3, 2, 1.5, 1, 0.5 Kb). All DNA samples were subjected to 1.0 % (w/v) agarose gel electrophoresis at 100 mA for 40 min (Section 2.2.2.1.4). Following electrophoresis, gels were soaked in a 10 μg/ml solution of EtBr for 10 min, washed with distilled water, and placed on a UV illuminator and a fluorescent image taken (Section 2.2.2.1.5).

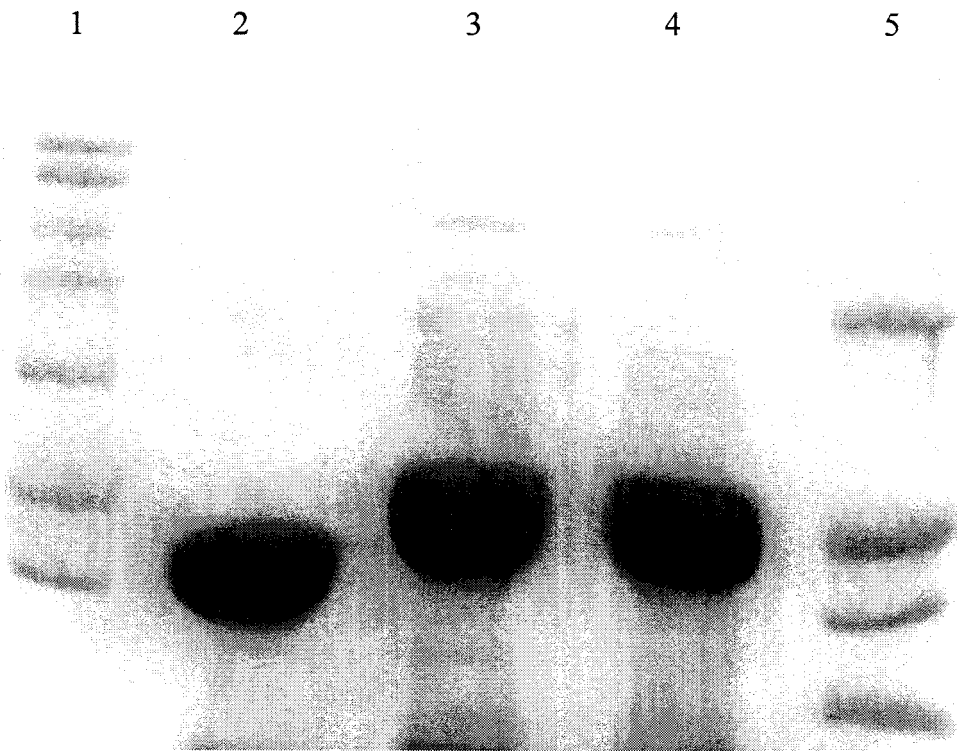


Figure 5.4: 12 % SDS-PAGE of HylP1, HylP2 and HylP3 purification.
HylP1, HylP2 and HylP3 (native recombinant) were purified from *E. coli* cells, harbouring plasmids of HylP1, HylP2 and HylP3 encoding putative hyaluronidases, by IMAC using chelating sepharose™. Each recombinant protein possessed an N-terminal His_6 tag. A gradient elution technique was used with the FPLC® at a linear flow rate of 149.0 cm/h (Section 2.2.3.4.2). Lane 1: High molecular weight (HMW) markers (205, 116, 97, 84, 66, 55, 45 kDa); lane 2: 20 µl of 1:10 HylP1 (38.4 kDa); lane 3: 20 µl of 1:10 HylP2 (42.0 kDa); lane 4: 20 µl of 1:10 HylP3 (41.8 kDa); lane 5: Low molecular weight (LMW) markers (66, 55, 45, 36 kDa). Proteins were visualised by staining with Coomassie Blue and destained (Section 2.2.3.1.1).

5.2.2 Expression and purification of Seleno-methionine HylP1, HylP2 and HylP3

For crystallisation and 3D structure determination, Se-Met containing HylP1, HylP2 and HylP3 were produced. HylP1, HylP2 and HylP3 plasmid DNA coding for a His_6 tag at the N-terminus, previously cloned (Section 5.1.1), was chemically transformed into *E.coli* B834 chemically competent cells (DE3; Section 2.2.2.1.13.1) and the recombinant *E. coli* strain was grown at 37 °C until an OD_{550} of ~ 0.2 was obtained. The cells were harvested and washed, as described in section 2.2.1.2.4, and used to inoculate the Se-Met media (Section 2.1.3.1.4). The recombinant *E. coli* strain was then grown at 37 °C to mid-exponential phase before induction with 1 mM IPTG (final concentration) and the cultures incubated at 30 °C overnight (Section 2.2.1.2.4). Cells were harvested by centrifugation and cell free extracts were prepared as described in section 2.2.3.3.

The proteins were purified from cell free extracts by IMAC under non-denaturing conditions, using a chelating sepharose column and an automated gradient elution technique (Section 2.2.3.4.2). The eluted proteins were dropped into tubes containing 10 mM DTT (final concentration) and the proteins were checked for purity by SDS-PAGE (Section 2.2.3.1), their concentrations determined by Bradford's assay (Section 2.2.3.7.2) and subjected to crystallisation studies (Section 2.2.3.8).

5.3 Kinetic analysis of HylP1, HylP2 and HylP3

5.3.1 Michaelis-Menten parameters for hyaluronate hydrolysis

To investigate the activity of the N-terminally tagged HylP1, HylP2 and HylP3 against hyaluronate, the enzymes were assayed for reducing sugar production with varying concentrations of the dialysed substrate sodium hyaluronate (Section 2.1.10.2). For all absolute kinetic analysis, 4 to 5 substrate concentrations (determined empirically and spanning K_M) in the limiting range were chosen and reactions were performed in triplicate. In order to ascertain if these data reflected true Michaelis-Menten kinetics, Lineweaver-Burk plots from three independent data sets were constructed as illustrated in Figures 5.5-5.7 and used to determine values of K_M, k_{cat} and k_{cat} / K_M (Table 5.1).

Enzyme	K_M (mg ml^{-1})	k_{cat} (s^{-1})	k_{cat} /K_M (s^{-1} ml mg^{-1})
HylP1	0.90	1390.90	1545.44
HylP2	2.07	742.01	358.80
HylP3	4.35	1253.04	287.79

Table 5.1: Michaelis-Menten kinetic parameters for the hydrolysis of dialysed sodium hyaluronate by HylP1, HylP2 and HylP3.
The reactions were performed in triplicate at 37 °C, in 20 mM Tris, pH 7.5 (HylP1), 20 mM MES, pH 6.0 (HylP2) and 20 mM sodium citrate buffer, pH 5.5 (HylP3), as described in section 2.2.5.1.4. BSA was included in the reaction at a final concentration of 1 mg ml^{-1}, except when HylP3 was assayed as precipitation occured.

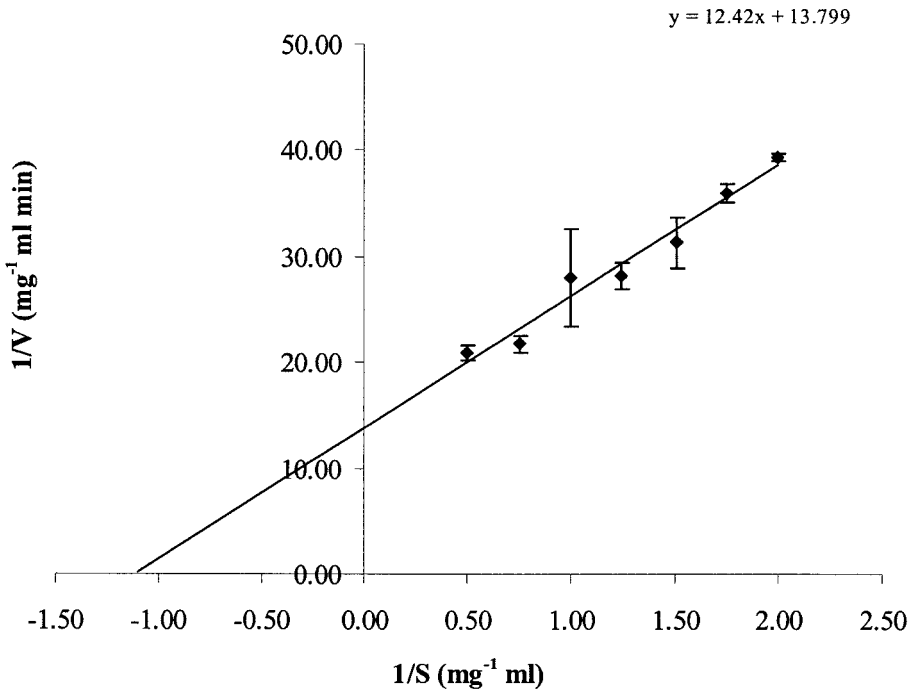


Figure 5.5: Lineweaver-Burk plot for HylP1 against dialysed sodium hyaluronate. These data were performed in triplicate with error bars representing the standard deviation from the mean. The reactions were performed at 37 °C, in 20 mM Tris buffer, pH 7.5, as described in section 2.2.5.1.4. BSA was included in the reaction at a final concentration of 1 mg ml^{-1}.

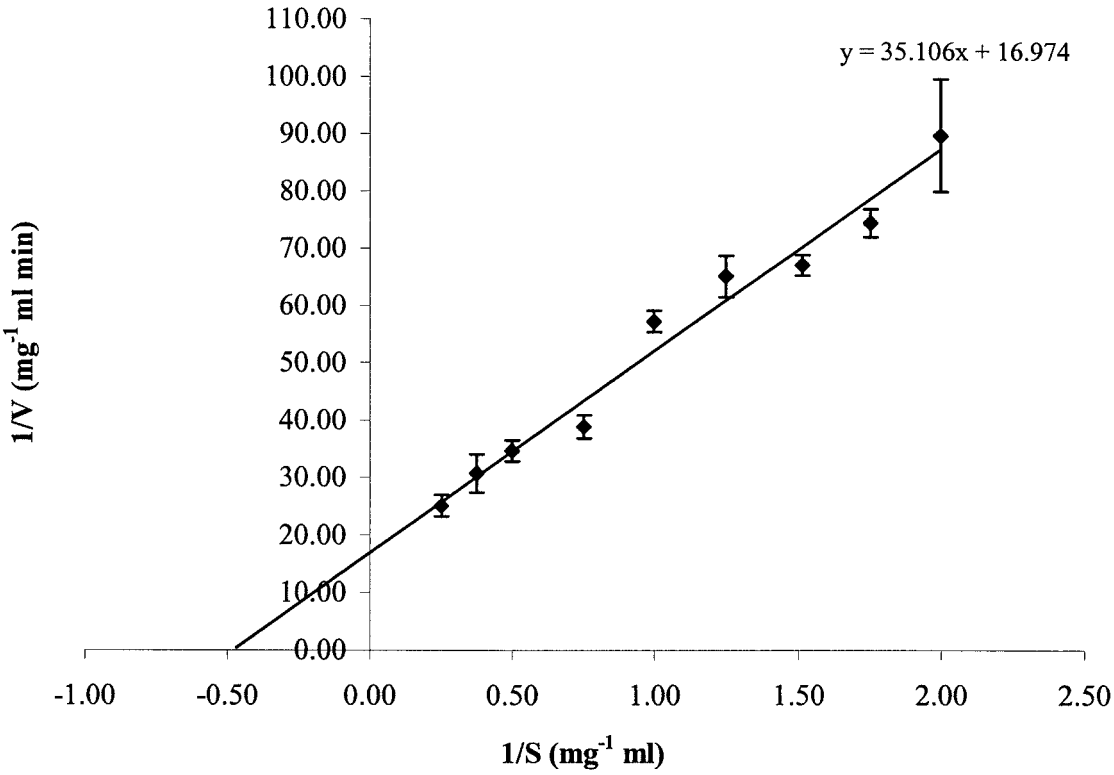


Figure 5.6: Lineweaver-Burk plot for HylP2 against dialysed sodium hyaluronate. These data were performed in triplicate with error bars representing the standard deviation from the mean. The reactions were performed at 37 °C, in 20 mM MES buffer, pH 6.0, as described in section 2.2.5.1.4. BSA was included in the reaction at a final concentration of 1 mg ml^{-1}.

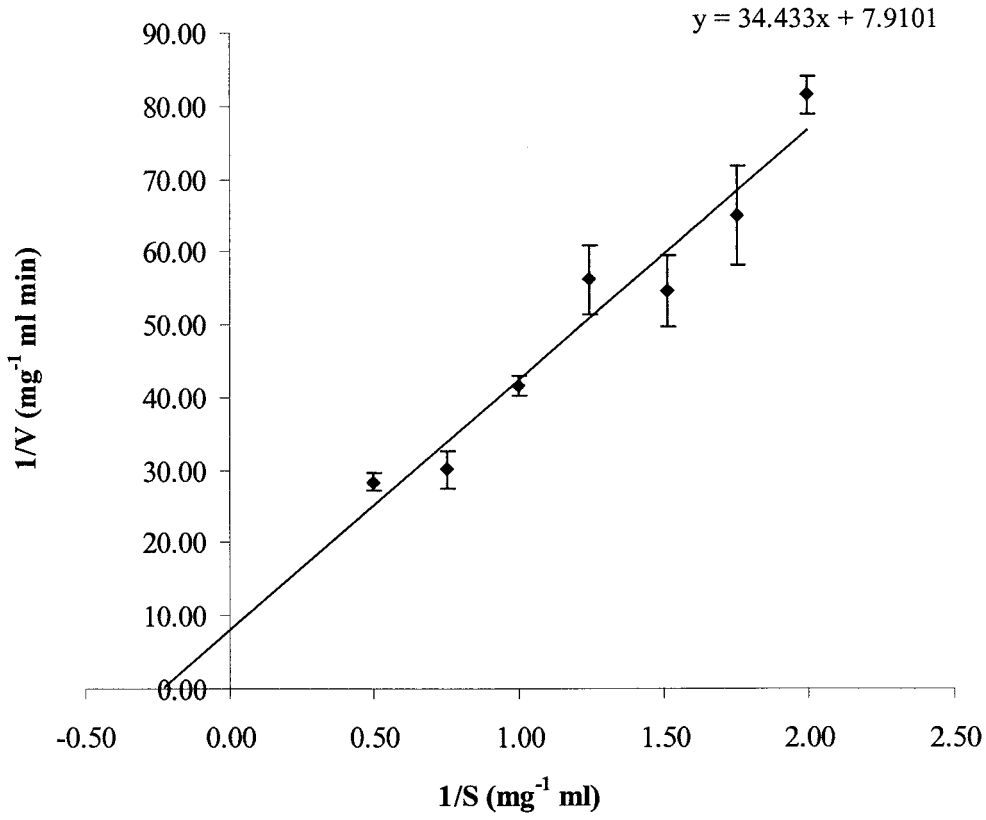


Figure 5.7: Lineweaver-Burk plot for HylP3 against dialysed sodium hyaluronate. These data were performed in triplicate with error bars representing the standard deviation from the mean. The reactions were performed at 37 °C, in 20 mM sodium citrate buffer, pH 5.5, as described in section 2.2.5.1.4. No BSA was added to the reaction as precipitation occurred.

5.3.2 Calculation of K_M, k_{cat} and k_{cat} / K_M from raw data

The calculation of k_{cat} (s^{-1}) required the use of three constants. Firstly, the relative molecular mass (RMM) of glucose ($C_6H_{12}O_6$) was calculated to be 180.0. Secondly, the number of molecules in one mole (Avagadro's number; 6.022 x 10^{23}) was used to calculate the number of molecules of product produced per second. Finally, for the conversion of enzyme weight in Daltons to weight in grams, the constant for the weight of one Dalton in grams was used: 1.661 x 10^{-24} g.

K_M and k_{cat} were then calculated from the Lineweaver-Burk plot as follows:

K_M (mg ml^{-1}) = -1 / (X intercept on Lineweaver-Burk plot)

V_{max} (U mg^{-1}) = (1 / y) / 10^3 / RMM Glucose x 10^6 / (Total protein added in mg)

k_{cat} (s^{-1}) = V_{max} / 10^6 / 60 x (6.022 x 10^{23}) / W

Where,

1 / y = gradient of line when x = 0

U = μmol of product produced per minute

W = Number of enzyme molecules per mg of protein: 0.001 g / (weight of 1 enzyme in Daltons x (1.661 x 10^{-24})).

K_M was calculated simply by dividing –1 by the X intercept in the Lineweaver-Burk plot to give K_M in mg ml^{-1}.

k_{cat} is calculated by first obtaining a value for V_{max} in μmoles of product produced per minute per mg (U mg^{-1}) and was calculated as follows: if x = 0, then y = y intercept, then 1 / y = mg of product produced per min. This value is then divided by 10^3 to convert into g of product produced per min. This value is then divided by the RMM of glucose (180.0 g L^{-1}) to give moles of product produced per min, which is subsequently multiplied by 10^6 to convert into μmoles of product

produced per min and is defined as U. The total reaction volume of 900 µl was taken into account and U was multiplied by 0.9. U is then divided by the total amount of enzyme added to the assay (mg). This gives the value for V_{max}, which is: Units of enzyme activity per mg (U mg^{-1}). V_{max} is then converted from µmoles of product produced per min per mg, into moles of product produced per minute per mg, by dividing V_{max} by 10^6. Moles of product produced per minute per mg, is then converted into moles of product produced per second per mg by dividing by 60. Moles of product produced per second per mg is then converted into molecules of product produced per second per mg, by multiplying by Avagadros number (6.022×10^{23}). Finally, the number of enzyme molecules present in 1 mg (W) is calculated. Firstly, the weight of one enzyme in grams is calculated by multiplying the enzyme weight in Daltons by the weight of 1 Dalton in grams (1.661×10^{-24}). The weight of 1 mg in grams, 0.001g, is then divided by the weight of one enzyme in grams to yield the number of enzyme molecules present in 1 mg (W). The number of molecules of product produced per second is then divided by the number of enzyme molecules present in 1 mg to yield the number of molecules of product produced per second per enzyme molecule, known as k_{cat}.

Specificity constants (s^{-1} ml mg^{-1}) were determined simply by dividing k_{cat} by K_M.

5.4 Biochemical and biophysical parameters of HylP1, HylP2 and HylP3

5.4.1 Determination of pH optimum

The N-terminally tagged HylP1, HylP2 and HylP3 proteins were shown to have an optimum pH of 6.5, 6.0 and 5.5, respectively, against dialysed sodium hyaluronate (2 mg ml^{-1} final concentration; section 2.2.5.1.5). A range of sodium citrate buffers were used to determine pH optimum of each enzyme. Assays were carried out at 37 °C and the release of reducing sugar was measured over time. Figure 5.8 shows the effect of pH on enzyme activity and each curve was constructed from three independent data sets.

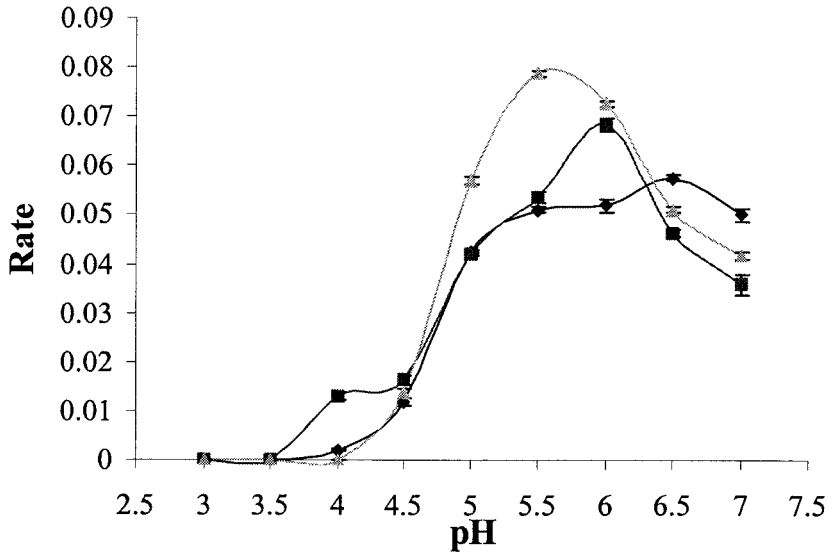


Figure 5.8: The effect of pH on the rate of enzyme activity against dialysed sodium hyaluronate.
The reactions were performed in triplicate, with error bars representing the standard deviation from the mean, at 37 °C and at final substrate concentration of 2 mg ml^{-1}. Sodium citrate buffers (20mM final concentration) was used in the pH range 3.0-7.0. No BSA was added to the reaction mix, as precipitation occurred in the presence of sodium citrate buffers, as described in section 2.2.5.1.5. Each point of the graph represents the mean for HylP1 (♦), HylP2 (■) and HylP3 (▲).

5.4.2 Determination of temperature and thermostability

At 37 °C, HylP1, HylP2 and HylP3 displayed the highest amount of activity against the dialysed substrate sodium hyaluronate (Section 2.2.5.1.6), when added chilled to a pre-warmed reaction mix, as shown in Figure 5.9. At temperatures above and below 37 °C, all three enzymes exhibited significantly reduced activity and complete inactivation between 67 °C and 77 °C.

Upon pre-incubation for 20 min at various temperatures prior to assay (Section 2.2.5.1.6), the enzymes HylP1, HylP2 and HylP3 retained the highest activity at 27 °C and were completely inactivated above 47 °C, as shown in Figure 5.10.

The effect of temperature on enzyme activity and stability was determined from three independent data sets, with the release of reducing sugar content being measured over time.

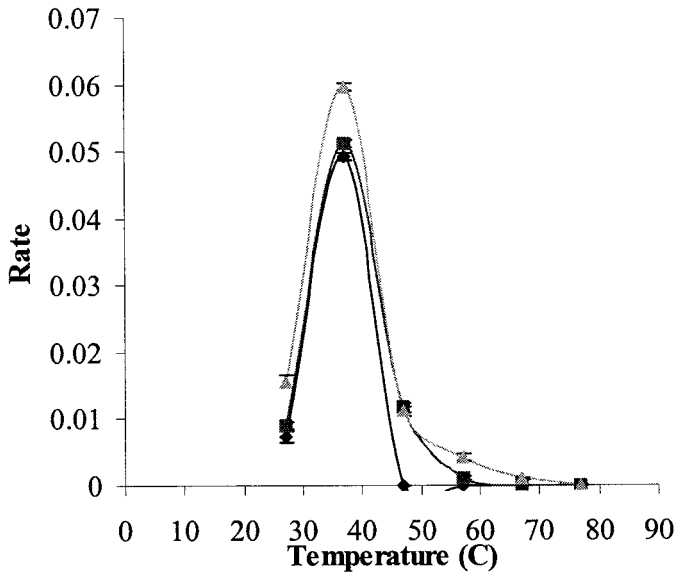


Figure 5.9: The effect of temperature on the rate of enzyme activity against dialysed sodium hyaluronate.
The reactions were performed in triplicate, with error bars representing the standard deviation from the mean, at 27, 37, 47, 57, 67, and 77 °C, at a final substrate concentration of 2 mg ml^{-1}. 20 mM Tris, pH 7.5 (HylP1), 20 mM MES, pH 6.0 (HylP2) and 20 mM sodium citrate buffer, pH 5.5 (HylP3), were used, as described in section 2.2.5.1.6. BSA was included in the reaction at a final concentration of 1 mg/ml, except when HylP3 was assayed as precipitation occured. Each point of the graph represents the mean for HylP1 (♦), HylP2 (■) and HylP3 (▲).

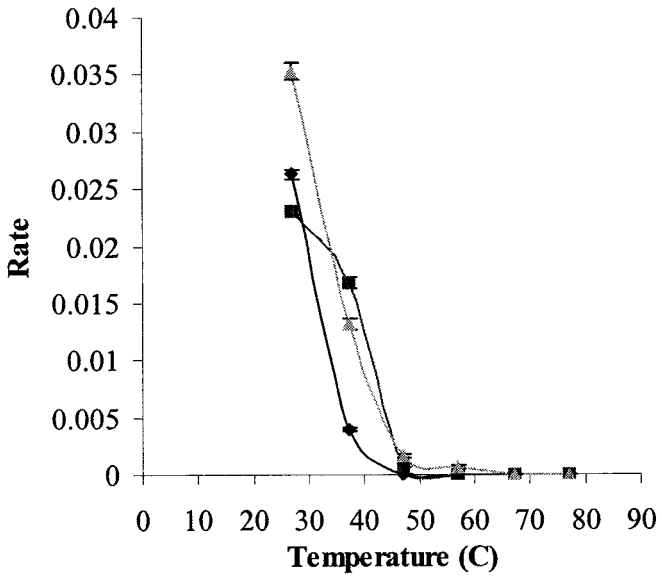
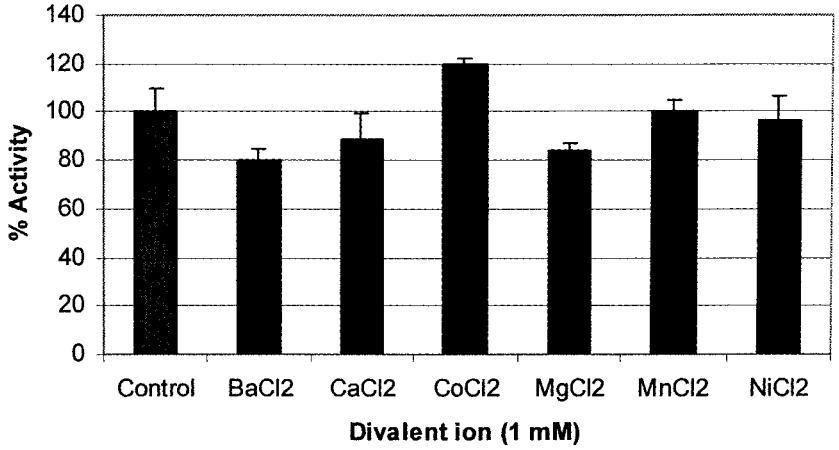


Figure 5.10: The effect of temperature on enzyme stability as measured by rate of activity against dialysed sodium hyaluronate.
The reactions were performed in triplicate, with error bars representing the standard deviation from the mean, at 37°C and at a final substrate concentration of 2 mg ml^{-1}, after pre-incubation of each enzyme for 20 min at 27, 37, 47, 57, 67, and 77 °C. 20 mM Tris, pH 7.5 (HylP1), 20 mM MES, pH 6.0 (HylP2) and 20 mM sodium citrate buffer, pH 5.5 (HylP3), were used, as described in section 2.2.5.1.6. BSA was included in the reaction at a final concentration of 1 mg/ml, except when HylP3 was assayed as precipitation occured. Each point of the graph represents the mean for HylP1 (♦), HylP2 (■) and HylP3 (▲).

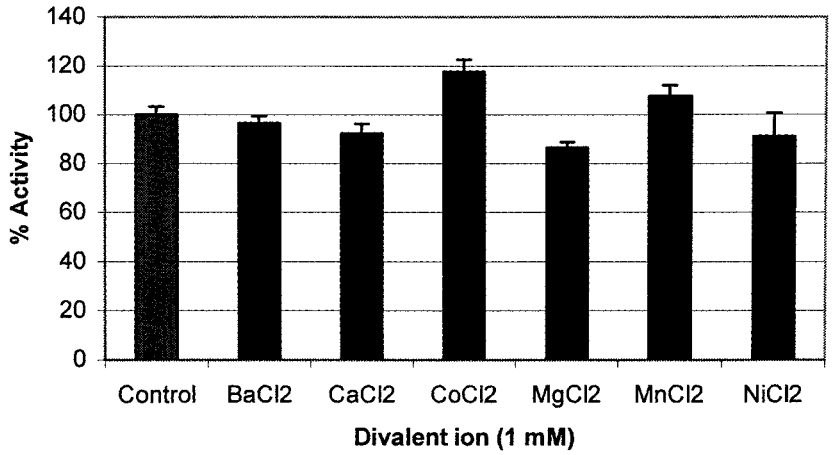
5.4.3 Determination of divalent ion requirement

When HylP1, HylP2 and HylP3 were tested with a range of divalent cations (Section 2.2.5.1.7; 1 mM final concentration), it was observed that none of the enymes had an absolute requirement for any of the ions tested, as the activity of each enzyme was not affected significantly. The data for the effects of divalent ions on enzyme activity is presented in Figure 5.11. Enzyme activity for HylP1 was shown to decrease slightly when $MgCl_2$ was added; HylP2 activity was slightly reduced in the presence of $BaCl_2$ and $MgCl_2$ and HylP3 showed the most significant decrease in enzyme activity when $MnCl_2$ was added. Optimal enzyme activity was observed when Co^{2+} ions were added to the reaction for HylP1, HylP2 and HylP3.

A



B



C

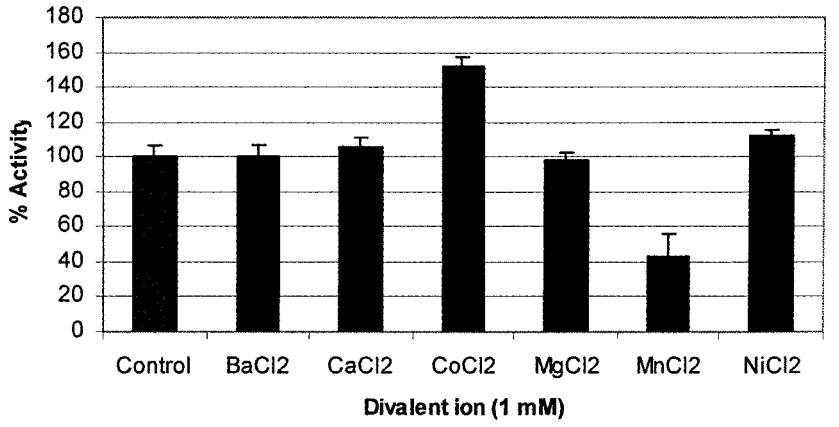


Figure 5.11: The effect of divalent cations on enzyme activity (A) HylP1, (B) HylP2 and (C) HylP3.

Reactions were performed in triplicate, with error bars representing the relative standard deviation from the mean, at 37 °C and at a final substrate concentration of 2 mg ml^{-1}, in 20 mM Tris, pH 7.5 (HylP1), 20 mM MES, pH 6.0 (HylP2) and 20 mM sodium citrate buffer, pH 5.5 (HylP3), as described in section 2.2.5.1.6. BSA was included in the reaction at a final concentration of 1 mg/ml, except when HylP3 was assayed as precipitation occured. Divalent cations (Ba^{2+}, Ca^{2+}, Co^{2+}, Mg^{2+}, Mn^{2+} and Ni^{2+}) were added to reactions at a final concentration of 1 mM. Each column represents the mean value for (A) HylP1, (B) HylP2 and (C) HylP3.

5.5 Mode of enzymatic action of HylP1, HylP2 and HylP3 enzymes by PAGE analysis

To determine whether all three enzymes share a common mode of action against sodium hyaluronate, the release of reducing sugar was followed with time as essentially described in section 2.2.5.1.3. Partial digestion of sodium hyaluronate with HylP1, HylP2, and HylP3 yields a homologous series of oligosaccharides, as demonstrated by PAGE (Section 2.2.5.2). Hyaluronate samples digested to differing extents are separated into a ladder-like series of bands. Since all of the hyaluronate fragments generated by enzymatic digestion have the same charge-to-mass ratio, separation occurs on the basis of fragment size, with the smallest species migrating most rapidly. Figure 5.12 shows typical electrophoretic patterns for an arbitrary hyaluronate oligosaccharide mixture following digestion with HylP1, HylP2 and HylP3. Upon analysis of the digestion products of HylP1, HylP2, and HylP3, each enzyme appears to display a random endolytic mode of cleavage on sodium hyaluronate yielding a range of oligosaccharides. Undigested high molecular weight sodium hyaluronate shows only limited migration into the 15.5 % polyacrylamide gel matrix (Figure 5.12).

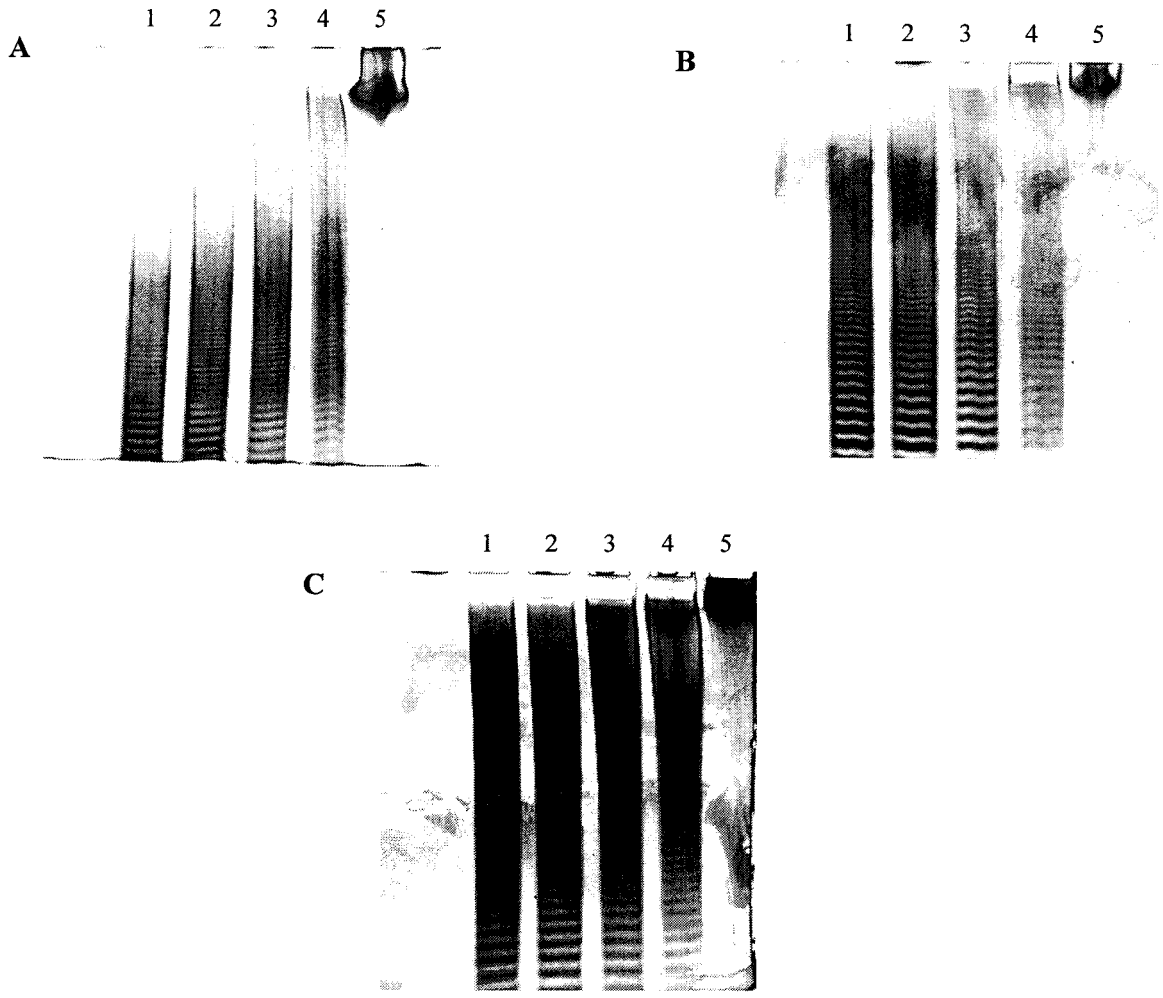


Figure 5.12: PAGE-stained patterns of HA samples digested with (A) HylP1, (B) HylP2 and (C) HylP3 over time.
Sodium hyaluronate (2 mg ml^{-1} final concentration) in 20 mM HEPES, pH 7.0, was digested with HylP1 (0.17 μg), HylP2 (0.28 μg) or HylP3 (0.35 μg) at various time points at 37 °C. Digests were terminated by boiling for 5 min. Digests were separated on a 15.5 % polyacrylamide gel under non-denaturing conditions in TBE buffer (Section 2.2.5.2). Gels were electrophoresed at +12 °C at 250 V for 20 min, 500 V for 10 min and 450 V for 5min, or until the Bromophenol blue tracking dye (Section 2.2.1.9.3.6) reached within ~1 cm of the bottom of the gel. (A) HylP1: lane 1: 60 min, lane 2: 40 min, lane 3: 20 min, lane 4: 10 min, and lane 5: 0 time. (B) HylP2: lane 1: 60 min, lane 2: 40 min, lane 3: 20 min, lane 4: 10 min, and lane 5: 0 time. (C) HylP3: lane 1: 60 min, lane 2: 40 min, lane 3: 20 min, lane 4: 10 min, and lane 5: 0 time. Oligosaccharides were stained with a combined Alcian Blue and Silver staining method, as detailed in Section 2.2.5.2.

5.6 Collagen-binding assay

A CLUSTAL W (1.82; Figure 5.2) multiple sequence alignment of a representative sample of ORFs from GH family 69 showed that three putative genes encoded proteins with internal repeats of the motif sequence Gly-X-Y. These amino acid triplet repeats resemble the characteristic repeating sequences found in collagen. Sequence analysis showed ORFs HylP2 and HylP3 from *S. pyogenes* SF370 to contain this collagen repeat, the function of which is not as yet known.

A crude binding assay was carried out to determine whether the N-terminally tagged HylP2 and HylP3 proteins could bind to insoluble collagen through their internal collagen-like repeat (Section 2.2.5.6). The same assay was also carried out in the presence of the N-terminally tagged HylP1 protein and BSA, as negative non-binding controls. Each protein (HylP1, HylP2, HylP3 and the BSA control) could be detected in the washes after binding, indicating that there was no binding (data not shown).

5.7 Crystallisation of HylP1, HylP2, and HylP3 enzymes

5.7.1 Native recombinant crystallisation of HylP1, HylP2 and HylP3

HylP1, HylP2, and HylP3 N-terminally tagged proteins were purified to apparent homogeneity (as judged by SDS-PAGE; section 2.2.3.1; Figure 5.4) and buffer exchanged into 10 mM HEPES, 50 mM $CaCl_2$, pH 7.4. The proteins were then concentrated to between 10 – 20 mg ml^{-1} for crystallisation studies. Crystals of HylP1, HylP2, and HylP3 were grown by vapour phase diffusion using the hanging-drop method (Section 2.2.3.8.2). HylP1 crystals were obtained in HSI 33 (4 M NaF; Appendix C), HylP2 in HSII 1 (2 M NaCl, 10 % (w/v) PEG 6000; Appendix C) and CSSI 22 (0.2 M KBr, 8 % (w/v) PEG 20K, 8 % (w/v) PEG 550mme; Appendix C) and HylP3 in HSI 6 (0.2 M $MgCl_2$, 0.1 M Tris-HCl pH 8.5, 30 % (w/v) PEG 4000; Appendix C). Optimisation studies for HylP1, HylP2, and HylP3 were then carried out around these conditions with the addition of different cryoprotectants in the drops. For HylP1, equal amounts of protein and

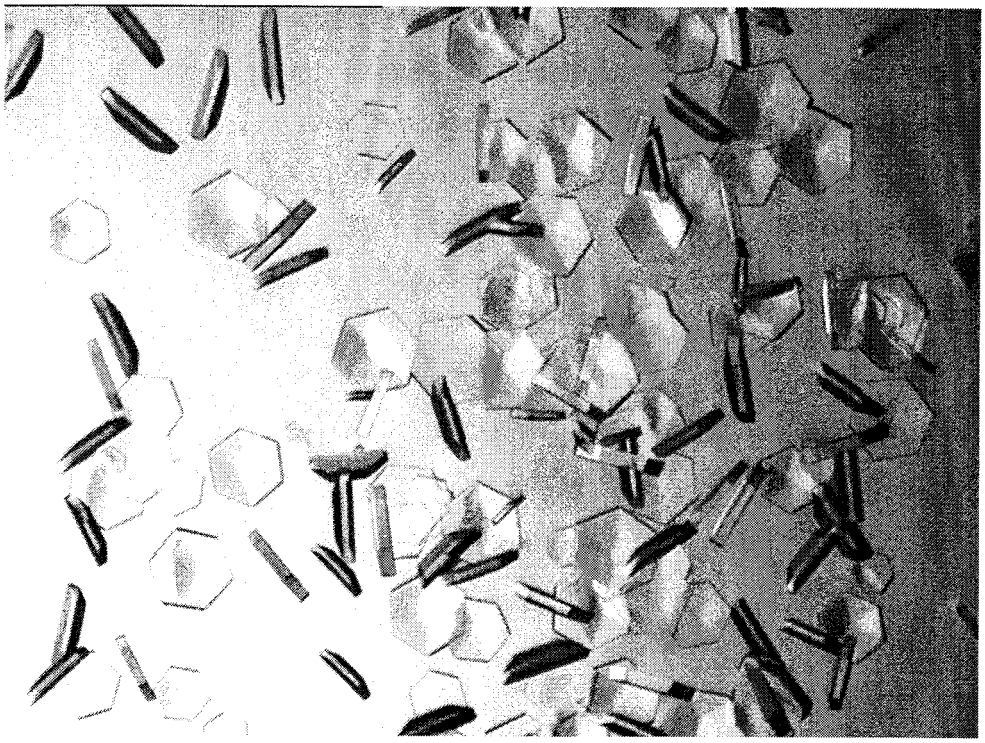


Figure 5.13: Crystals of native recombinant HylP1 from *S. pyogenes* SF370. Crystals were grown by the hanging-drop method (Section 2.2.3.8.2). Equal amounts of protein and reservoir solution (3.5 M Sodium formate, final concentration) were mixed and MPD (15 % (v/v) final) was added to the drop. Hexagonal crystals grew after a few days at 22 °C.

reservoir solution, containing 3.5 M sodium formate, were mixed. MPD (15 % (v/v) final) was also added to the drop. Hexagonal crystals grew after a few days at 22 °C (Figure 5.13). A single crystal was removed from the drop using a rayon fibre loop and transferred into a cryoprotectant solution consisting of the crystal growth buffer supplemented with 15 % (v/v, final) MPD, prior to freezing the crystals in liquid nitrogen. A rayon fibre loop was used to mount the crystal and an image collected using a Rigaku X-ray generator and a MAR 30 cm image plate at YSBL. Crystals that diffracted beyond 2.8 Å were saved and stored in liquid nitrogen for further analysis at the ESRF at Grenoble. The structural data obtained to date for HylP1 and HylP3 is given in Table 5.2. All data collection and analysis was performed by Dr E. Taylor, Dr Johan Turkenburg, Prof. Eleanor Dodson and Prof. Gideon Davies at YSBL.

5.7.2 Se-Met crystallisation of HylP1

As there were no other related structural modules available for molecular replacement, anomalous data was required to solve one of the HylP structures. Due to the high degree of homology between all three structures (CAZy) it was likely that the structure solution of one of the HylP proteins would provide a molecular replacement module to the remaining two.

There are two general ways of obtaining anomalous data. The incorporation of seleno-methionine into the protein at translation, normally achieved with the use of a methionine auxotroph and labelled seleno-methionine used in conjunction with a minimal media, or the use of a heavy metal soak. This is where heavy metal ions such a uranium, lead or mercury are soaked into the crystals, these bind to the protein specifically and uniformly. Both methods result in regular points of high electron density throughout the lattice. The difference between a native data set and one with an anomalous signal may be used to calculate the phases and solve the structure. Where only a single wavelength is used this is known as a single anomalous difference or SAD experiment. Where a tuneable radiation source is available, the anomalous signal may be optimised by using multiple wavelengths, which is known as a MAD experiment. MAD data sets are collected at the peak, inflection and a remote from a single crystal. Comparisons between these three sets, gives a more accurate calculation of phases and position of the

heavy atom. The HylP1 structure was solved using phase data obtained from a MAD experiment using incorporated seleno-methionine and a SAD experiment using a uranium acetate soak. These phases were then applied to a higher resolution native data set, and used to calculate the electron density. The scaling details for the solution of HylP1 may be seen in Table 5.2. DTT was maintained in the purification buffers to prevent the oxidation of the selenium (oxidised selenium has a more diffuse absorption edge).

The pET-28a vector harbouring the expressed inserts for HylP1, HylP2, and HylP3 were transformed into *E. coli* B834 (DE3; Section 2.2.2.1.13.1), an expression strain disabled with regards to methionine metabolism. The HylP1, HylP2, and HylP3 proteins were expressed, as described in section 2.2.1.2.4, in Se-Met minimal media (Section 2.1.3.1.4). HylP1, HylP2, and HylP3 N-terminally tagged proteins were purified to apparent homogeneity (as judged by SDS-PAGE; Section 2.2.3.1) and buffer exchanged into 10 mM HEPES, 50 mM $CaCl_2$, 1 mM DTT, pH 7.4. The proteins were then concentrated to between 10 – 20 mg ml^{-1} for crystallisation studies. The Se-Met-substituted protein did not produce crystals under the same conditions as the native protein and also in the presence of 1 mM DTT.

5.7.3 Seeding of HylP1 native and Se-Met forms

The technique of seeding crystals was used in order to try and encourage the Se-Met-substituted HylP1 protein to grow. This essentially involves adding microseeds of the native recombinant HylP1 crystals into the drop with both forms of the protein to help promote growth. Crystals of HylP1 in both native recombinant and Se-Met forms were grown from microseeds in 3.5 M sodium formate (Section 2.2.3.8.3) over a period of a few days and a few weeks respectively at 22 °C. Single crystals were then flash-cooled in liquid nitrogen and transferred into a liquid nitrogen cryo stream using a rayon fibre loop to determine diffraction quality.

5.7.4 Heavy metal derivatives of HylP1

This technique was employed to try and obtain better phasing power at lower resolutions in the absence of Se-Met substituted HylP1 crystals. The Ta_6Br_{12}

cluster compound was allowed to diffuse into the HylP1 native hexagonal crystals, as described in section 2.2.3.8.4, and mounted onto the goniometer in order to determine diffraction quality. The crystals were then preserved in liquid nitrogen and transported to the ESRF at Grenoble, where it was remounted on the single-axis goniometer for data collection. The data was collected by Dr E. Taylor, YSBL.

5.8 Diffraction analysis of HylP1 and HylP3

Preliminary X-ray diffraction analysis revealed that crystals of HylP1 belong to space group H32 with unit cell dimensions a = 58.53 Å, b = 58.53 Å and c = 586.58 Å for native recombinant HylP1 and space group P622 and unit cell dimensions of a = 49.48 Å, b = 49.48 Å and c = 241.0 Å for native recombinant HylP3 crystals. There is one molecule of HylP1 or HylP3 in the asymmetric unit. The data collected at the ESRF on beamline ID-14.2 is presented in Table 5.2.

	HylP1 native	HylP1 MAD peak	HylP1 MAD remote	HylP1 MAD inflection	Uranium derivative	HylP3 Native
Space group	H32	H32	H32	H32	H32	P622
Unit cell (Å) *a*b*c*	58.5341*58.53 41*586.5852	58.8138*58.81 38*587.8541	58.7976*58.79 76*588.4427	58.7639*58.76 39*588.8071	58.7966*58.79 66*588.3383	49.4760*49.47 60*241.0
α β γ (^{0})	90*90*120	90*90*120	90*90*120	90*90*120	90*90*120	90*90*120
Resolution of data (outer shell), Å	195.52-1.80 (1.90-1.80)	43.44-2.50 (2.64-2.50)	50.64-2.50 (2.64-2.50)	196.27-2.50 (2.64-2.50)	196.113-2.50 (2.64-2.50)	1.97-1.9
[a]*R* merge (outer shell)	0.097 (0.259)	0.062 (0.145)	0.064 (0.223)	0.055 (0.153)	0.086 (0.179)	0.261
Mean I/σI (outer shell)	3.9 (2.6)	7.0 (4.6)	55.5 (3.3)	7.5 (4.7)	5.2 (3.3)	2.9
Completeness (outer shell), %	98.8 (98.8)	99.9 (99.9)	99.9 (99.9)	99.8 (99.8)	99.6 (99.6)	51.6
Multiplicity (outer shell), %	5.5 (4.1)	5.2 (5.3)	5.7 (5.7)	5.0 (5.0)	9.6 (7.0)	1.8
No unique reflections	2157662	1178842	1438940	898782	1578676	13481
No rejections	1717	1127	449	756	339	3534
Crystallisation conditions	3.5M Sodium formate, 15% (v/v) MPD	3.5M Sodium formate, 15% (v/v) MPD	3.5M Sodium formate, 15% (v/v) MPD	3.5M Sodium formate, 15% (v/v) MPD	3.5M Sodium formate, 15% (v/v) MPD	30 % (v/v) PEG 2000, 0.1 M Tris, pH 8.5, 0.2 M $MgCl_2$, 5 % (v/v) glycerol
Beam line collected on (date)	ID 141 (15/11/03)	ID 144 (4/03/04)	ID 144 (4/03/04)	ID 144 (4/03/04)	ID 141 (15/11/03	ID 144 (15/11/03)
Wavelength	0.93400	0.97950	0.93930	0.97960	0.93400	0.93400

Table 5.2: Refinement and Structure Quality Statistics for the S. pyogenes SF370 HylP1 structure and preliminary X-ray diffraction data for HylP3. The unit cell dimensions, refinement and structure quality statistics for the data sets of HylP1 and HylP3. Number of molecules in the asymmetric unit for HylP1 and HylP3 is one. [a]$R_{merge} = \Sigma_{hkl}\Sigma_i\backslash I_{hkli} - \langle I_{hkl}\rangle / \Sigma_{hkl}\Sigma_i\backslash I_{hkli}$. All data within this table was collected and analysed by Dr Edward Taylor, in collaboration with YSBL.

5.9 Three-dimensional structure of HylP1

The 3D structure of HylP1 was solved at a resolution of 1.8 Å. The 3D structure (Figure 5.14), extending from residues 7 to 337 (LRVQ to LILKK from sequence; Figure 5.2), is composed of three monomeric strands, comprising α- and β-helical strands, that intertwine to form a trimer. Using molecular replacement, the structures of HylP2 and HylP3 have been determined and the structures are truncated forms of HylP1.

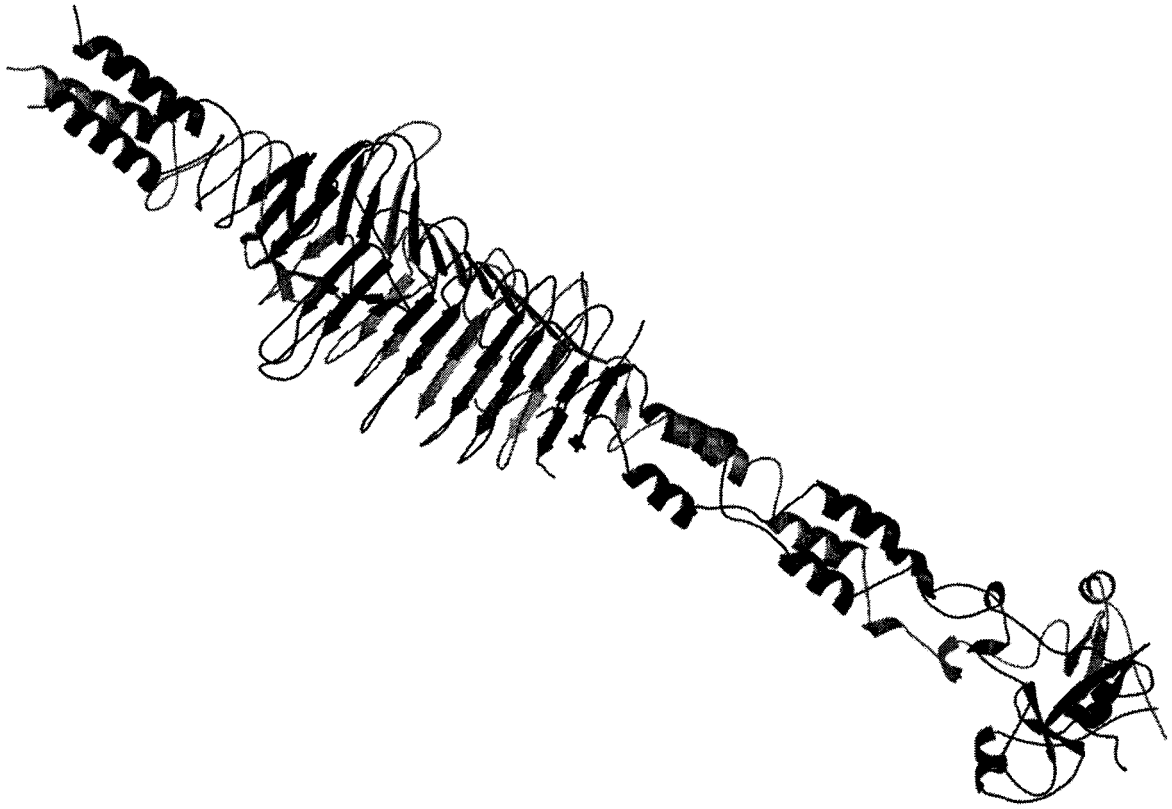


Figure 5.14: Structure of *S. pyogenes* SF370 HylP1 hyaluronidase.
Ribbon diagram of the side view of the trimer. There is one molecule of HylP1 in the asymmetric unit. The three HylP1 monomers are coloured red, green and blue.

Figure 5.15: Structure of *S. pyogenes* SF370 HylP1 hyaluronidase.
(A) Ribbon diagram of the C-terminal end of the trimer. (B) Ribbon diagram of the N-terminal end of the trimer. The three HylP1 monomers are coloured red, green and blue.

Biochemical characterisation and crystallisation studies of three hyaluronidases from GH69.

6 Discussion

6.1 Introduction

The primary aim of the work undertaken in this chapter was to biochemically characterise three structurally related putative hyaluronidases from *Streptococcus pyogenes* SF370 and to determine their three-dimensional (3D) structures. A reducing sugar assay was used to determine biochemical parameters for HylP1, HylP2 and HylP3 and polyacrylamide gel electrophoresis (PAGE) analysis was undertaken to identify the mode of action of each enzyme. Moreover, crystallisation studies of all three enzymes resulted in the 3D structure determination of HylP1 and truncated versions of HylP2 and HylP3.

6.2 Glycoside hydrolases with hyaluronidase activity

Glycoside hydrolases (GHs) are enzymes that hydrolyse the glycosidic bond found in polysaccharides, oligosaccharides, glycoproteins, glycosaminoglycans (GAGs) and proteoglycans, and have been classified into families based on primary structure similiarity (Henrissat, 1991; Henrissat and Bairoch, 1996; Coutinho and Henrissat, 1999).

Hyaluronidases are widespread in nature and are found in mammals, insects, leeches and bacteria (Meyer, 1971; Kreil, 1995; Frost *et al.*, 1996). The substrate for these enzymes, hyaluronan (hyaluronic acid; HA), is becoming increasingly important and is now recognised as a major participant in processes such as cell motility, wound healing, embryogenesis and has been implicated in cancer progression (Frost *et al.*, 1996). Also, hyaluronidase is associated with mechanisms of entry and spread, for example, as a virulence factor for pathogenic bacteria, for tissue dissection in gas gangrene, as a means of treponema spread in syphilis and for penetration of skin and gut by parasites (Frost *et al.*, 1996; Csóka *et al.*, 1997; Menzel and Farr, 1998). Hyaluronidase also comprises a component

of the venom of a wide variety of organisms, including bees, wasps, hornets, fish, snakes and lizards (Frost *et al.*, 1996; Csóka *et al.*, 1997).

On the basis of sequence similarity, mammalian and insect hyaluronidases have been classified as belonging to GH family 56. Hyaluronidase from bee venom (*Apis mellifera*; Hya) specifically degrades HA in the extracellular matrix of skin (Markorić-Housley *et al.*, 2000). Hya shares as much as 30-40 % sequence identity with several mammalian enzymes, including the sperm PH-20 protein (Gmachl and Kreil, 1993; Lin *et al.*, 1994; Markorić-Housley *et al.*, 2000); and the human lysosomal enzymes Hyal-1 (Csóka *et al.*, 1997; Frost *et al.*, 1997; Triggs-Raine *et al.*, 1999) and Hyal-2 (Lepperdinger *et al.*, 1998). Sperm PH-20 hyaluronidase has been shown to be essential for fertilization (Lin *et al.*, 1994; Frost *et al.*, 1996; Cherr *et al.*, 2001) and the human lysosomal enzymes have been shown to be involved in HA turnover (Frost *et al.*, 1997; Lepperdinger *et al.*, 1998). The hyaluronidase isolated from frog kidney (*Xenopus laevis*; XKH1) also belongs to GH56 and is required for HA degradation in this organ (Reitinger *et al.*, 2001).

The hyaluronidases from *Streptococcus pyogenes* bacteriophage H4489A (*hylP* and *hylP2*) show a high degree of similarity and can be found in GH69. Interestingly, one major region of difference is the deletion (or addition) of a collagen-like Gly-X-Y region in *hylP* (Hynes and Ferretti, 1989; Hynes *et al.*, 1995; Hynes and Walton, 2000). However, this region is missing in *hylP2*, suggesting that these repeats are not important for the enzymatic activity of hyaluronidase (Hynes *et al.*, 1995).

The origin of this collagen-like region of DNA in the hyaluronidase genes is unknown; however, it is clearly widespread, since the sequence is present in other strains (Hynes *et al.*, 1995). This study has shown that the three open reading frames (ORFs) HylP1, HylP2 and HylP3 from *S. pyogenes* SF370 do indeed code for active hyaluronidases and belong to GH69. As with the bacteriophage hylauronidase *hylP*, HylP2 and HylP3 possess the motif sequence Gly-X-Y. HylP1, however, like *hylP2*, lacks this region and it is therefore possible to

speculate that this collagen-like domain is not important for enzyme activity, but may play some other role.

6.3 Biochemical characterisation of HylP1, HylP2 and HylP3

Successful cloning and expression of ORFs HylP1, HylP2 and HylP3 from *S. pyogenes* SF370 has shown that these proteins possess hyaluronidase activity. This observation prompts the question why does *S. pyogenes* SF370 possess three enzymes capable of degrading its own capsule? It is known that the bacterium *S. pyogenes* possesses a HA capsule to evade phagocytosis by the host, as it is antigenically similar to HA of host connective tissue (Cunningham, 2000; Bisno *et al.*, 2003). It can be postulated that *S. pyogenes* SF370 requires one hyaluronidase to digest the HA capsule surrounding the bacterium and two other hyaluronidases, which contain a collagen domain, to aid in the spread of the organism through tissue invasion and infection.

Quantitative analysis of the kinetics of HylP1, HylP2 and HylP3, after the release of oligosaccharides from hyaluronate degradation, yields k_{cat} and K_M for native recombinant HylP1 of 1390.90 s^{-1} and 0.90 mg ml^{-1}; HylP2 of 742.01 s^{-1} and 2.07 mg ml^{-1}; and HylP3 of 1253.04 s^{-1} and 4.35 mg ml^{-1}, respectively. HylP1 showed a high specificity constant (1545.44 s^{-1} ml mg^{-1}) on sodium hyaluronate compared to HylP2 (358.80 s^{-1} ml mg^{-1}) and HylP3 (287.79 s^{-1} ml mg^{-1}). However, since no other type of hyaluronate was used to assay HylP1, HylP2 and HylP3, no comparisons can be made. The K_M for HylP1 is comparable to the K_M of pig liver hyaluronidase (0.64 mg ml^{-1}; Joy *et al.*, 1985). Unlike some of the hyaluronate lyases that have been characterised (Baker *et al.*, 1997), HylP1, HylP2 and HylP3 were not active against the substrate chondroitin.

Interestingly, the optimum pH for HylP1, HylP2 and HylP3 varies slightly (6.5, 6.0 and 5.5, respectively; Figure 5.8) against the substrate sodium hyaluronate. The significance, if any, of this is as yet to be established, but it may be speculated that there is some variation in the mechanistic and structural features of all three enzymes that allows them to function at a different pH. Alternatively, it is possible that HylP2 and HylP3 are mechanistically very similar to HylP1, however, the tertiary structure of the enzymes are unable to withstand the slightly

higher pH tolerated by HylP1. This may also explain the diminished activity HylP2 and HylP3 display against sodium hyaluronate. It may also be that *S. pyogenes* SF370 requires one hyaluronidase to firstly degrade its own HA capsule and secondly, two other hyaluronidases to degrade host HA during tissue invasion, depending on the degree of infection, thus would require hyaluronidases that function at different pHs. In comparison, purified bovine testicular hyaluronidase has a pH optimum for the hydrolysis of potassium hyaluronate at 5.2 in the presence of NaCl, and at 6.0 in the absence of NaCl (Gorham *et al.*, 1975). Pig liver hyaluronidase activity was measured on potassium hyaluronate at a pH optimum of 4.0 in the presence of 0.1 M NaCl and a pH optimum of 3.4 in the presence of 0.3 M NaCl (Joy *et al.*, 1985). The activity of hyaluronidases from human serum and placenta was optimum between pH 3.4 and 4.0 (Fiszer-Szafarz *et al.*, 2000). Characterisation of human urinary hyaluronidase was found to have a pH optimum of 3.4. The presence of 0.1 M NaCl caused an increase in pH optimum to pH 3.8, and a decrease in overall activity (Csóka *et al.*, 1997).

Given that *S. pyogenes* SF370 is a known strict human pathogen (Ferretti *et al.*, 2001) and therefore unlikely to be exposed to temperatures greater than 37 °C, it is therefore understandable that HylP1, HylP2 and HylP3 displayed optimum catalytic performance at this temperature (Figure 5.9). Samples of HylP1, HylP2 and HylP3 were heated at various temperatures from 27 to 77 °C for 20 min and assayed for activity. Figure 5.10 shows that the thermal stability profiles for HylP1, HylP2 and HylP3 are nearly identical. The activity of these enzymes after pre-incubation at various temperatures showed that these enzymes are not very thermostable. The highest activity was observed after pre-incubation at 27 °C, however, all three enzymes had lost ~ 50 % activity compared to when HylP1, HylP2 and HylP3 were assayed at 37 °C without pre-incubation of the enzyme. Pre-incubation of HylP1, HylP2 and HylP3 at 37 °C produced a significant decrease in enzyme activity and at temperatures > 47 °C, all three enzymes were inactivated. In comparison, human serum and placenta hyaluronidases faired slightly better when thermal stability profiles were carried out. Samples were heated at various temperatures from 37 to 70 °C for 5 min and were shown to be stable up to 46 °C (Fiszer-Szafarz *et al.*, 2000).

In order to assay the effects of divalent ions on the activity of HylP1, HylP2 and HylP3, various divalent ions were added to the reaction at a final concentration of 1 mM to ensure any effects were observed. HylP1, HylP2 and HylP3 did not show any requirement for divalent ions (Figure 5.11). It should be noted, however, that although the enzymes did not shown any requirement for divalent ions during the degradation of hyaluronate, $CaCl_2$ was shown to be essential during the purification process. Elimination of $CaCl_2$ during the purification process resulted in precipitation of each protein. The role of Ca^{2+}, if any, is still to be established, but it maybe necessary for the stability of each enzyme. Further analysis of the HylP1 structure is therefore needed to see if a metal ion is present in the structure and at the current stage of refinement of HylP1, there appears to be a metal ion present.

6.4 Mode of enzymatic action of HylP1, HylP2 and HylP3

Oligomeric and polymeric fragments of GAGs can be separated by electrophoresis through a polyacrylamide matrix. A ladder-like series of bands is observed, in which bands correspond to species differing in chain length. HylP1, HylP2 and HylP3 display a 'random' endo depolymerisation pattern against sodium hyaluronate, with the generation of a range of oligosaccharides (Figure 5.12). PAGE analysis shows clearly this random endolytic mode of cleavage pattern, similar to that observed by Min and Cowman (1986) and Ikegami-Kawai and Takahashi (2002). In both studies, HA samples were digested to varying extents by bovine testicular hyaluronidase to yield a homologous series of oligosaccharides in a ladder-like series of bands.

Random endolytic cleavage suggests that HylP1, HylP2 and HylP3 are able to adopt a wide variety of substrate binding modes, leading to the gradual degradation of longer oligomers into shorter oligomers without resulting in a disproportionate accumulation of shorter products early in the reaction. Upon further analysis of the 3D structure of HylP1, and therefore HylP2 and HylP3, it may be found that the active site topology displayed by HylP1 is an 'open' cleft active site, which is known to be a common feature of random binding hydrolytic polysaccharide degraders (Davies and Henrissat, 1995).

6.5 3D structure of HylP1

The 3D structure of HylP1 from *S. pyogenes* SF370 was solved at a resolution of 1.8 Å and was shown to belong to space group H32. The 3D structure (Figures 5.14 and 5.15) is composed of three monomers, made up of α- and β-helical strands that intertwine to form a trimer. The topography of HylP1 reveals distinct similarities with the tail protein from bacteriophage T4 (Kanamaru *et al.*, 2002). HylP1 is a phage-associated hyaluronidase from *S. pyogenes* SF370 and it can be seen from the 3D structure to possess the phage tail characteristics from the bacteriophage T4. This resemblance allows speculation of the role HylP1 may play in *S. pyogenes* SF370. Assuming that HylP1 is a phage structural protein, how does HylP1 interact with the tail fibre proteins? This is analogous to the lysozyme domain containing cell-puncturing device of the recently solved 3D structure of bacteriophage T4 (Kanamaru *et al.*, 2002). Unlike bacteriophage T4 that has a lysozyme domain linked to the tail fibre protein, the hyaluronidase component of HylP1 is is an integral part of the tail fibre protein. It is possible that the phage tail from HylP1 is likely to adhere to the cell membrane of *S. pyogenes*, which will faciliate degradation of the HA capsule that surrounds the bacterium. Bacteriophage T4 has a very efficient mechanism for infecting cells (Kanamaru *et al.*, 2002). The key component of this process is the baseplate, located at the end of the phage tail, which is required to penetrate the outer cell membrane of *Escherichia coli*. The interaction of the baseplate tail fibres with the outer cell membrane disrupts the intermembrane peptidoglycan layer, promoting subsequent entry of phage DNA into the host (Kanamaru *et al.*, 2002). The 3D structure of bacteriophage T4 shows that it is a modular protein, which possesses lysozyme activity that is attributable to the attached lysozyme domain. This lysozyme domain is required to digest the cellular intermembrane peptidoglycan layer during phage tail attachment (Kanamaru *et al.*, 2002). It is therefore possible to speculate that HylP1 uses a similar process in *S. pyogenes* SF370. Unlike bacteriophage T4, HylP1 is not a modular protein, but it is possible to speculate that like bacteriophage T4, HylP1 uses its phage-like tail to anchor itself to the cell membrane of S. *pyogenes* to enable the hyaluronidase component to digest the surrounding HA capsule. Once the capsule is degraded, the bacterium is then free to eject enzymes into the host and cause infection. It may be possible that the

other two phage-encoded hyaluronidases (HylP2 and HylP3) expressed by *S. pyogenes* SF370 are used to facilitate spread of the organism through host tissues during infection by degrading the HA found in host connective tissue.

A number of questions arise with regard to the phage-associated HylP1, HylP2 and HylP3 hyaluronidases. Firstly, are HylP1, HylP2 and HylP3 involved in *S. pyogenes* infection? This could possibly be determined by examining expression levels of prophage-encoded putative virulence factors by growing a co-culture of *S. pyogenes* strain SF370 with epithelial pharyngeal cells, for example, to determine whether the prophage-encoding hyaluronidases HylP1, HylP2 and HylP3 are expressed during infection. Alternatively, it is likely that the *S. pyogenes* SF370 hyaluronidases will be maximally synthesised when the microbe is propagated in the presence of HA as the sole carbon source. Thus, by propagating *S. pyogenes* on this growth media and then separating the synthesised proteins using two-dimensional (2D) gel electrophoresis, i.e. by isoelectrofocusing followed by denaturing PAGE, it may be possible to identify expressed hyaluronidases. These proteins can then be purified from the gel, fragmented and their unique fragmentation profile determined using mass spectrometry. Secondly, nothing is known about the host and environmental signals or genetic regulatory features that control the expression of phage-associated hyaluronidases in *S. pyogenes* SF370. If HylP1, HylP2 and HylP3 are expressed during normal growth of *S. pyogenes* SF370, i.e. with bacterial promoter control, then are they secreted and how? The genome sequence of *S. pyogenes* SF370 (Ferretti *et al.*, 2001) has identified, due to sequence similarity searches, that this strain of *S. pyogenes* carries four ORFs that are likely to encode hyaluronidases. Apart from HylP1, HylP2 and HylP3 (GH69) that have been shown to be hyaluronidases (This thesis), the remaining hyaluronidase (SPy1600; GH84) annotated sequence has recently been shown to be an *N*-acetyl-β-glucosaminidase (Personal communication, Mr. W.L. Sheldon). Therefore, it is likely that the remaining three hyaluronidases are expressed, but what secretion system does *S. pyogenes* SF370 use to excise these enzymes extracellulary during infection? Finally, if HylP1, HylP2 and HylP3 are not expressed as discrete ORFs

in *S. pyogenes* SF370, then are they expressed only in association with the phage particle i.e. during induction of the lytic cycle?

6.6 Role of hyaluronidases

Hyaluronidase from bee venom has been shown to facilitate the process of degrading the skin extracellular matrix during the penetration of venom constituents into the body (Kemeny *et al.*, 1984; Gmachl and Kreil, 1993; Markorić-Housley *et al.*, 2000). Likewise, degradation of HA in the extracellular matrix of tissues is presumed to be the key event in the enzyme mediated spreading process during snakes' envenomation (Girish *et al.*, 2002).

The sperm protein PH-20 has hyaluronidase activity and has been shown to be present on the plasma membrane of human, monkey, mouse and guinea pig sperm (Gmachl *et al.*, 1993; Lin *et al.*, 1993, Lin *et al.*, 1994; Arming *et al.*, 1997; Primakoff *et al.*, 1998; Li *et al.*, 2002). The function of sperm protein PH-20 is to use its hyaluronidase activity to hydrolyse the HA-rich cumulus matrix that surrounds the egg during sperm penetration (Lin *et al.*, 1994; Li *et al.*, 2002). The hyaluronidase from *X. laevis* is exclusively expressed in the cells of the adult frog kidney where it either may be involved in the reorganisation of the extracellular architecture or in supporting physiological demands for proper renal functions (Reitinger *et al.*, 2001).

HA is the sole component of the capsule of group A streptococci (Hynes *et al.*, 1995; Cunningham, 2000; Bisno *et al.*, 2003). The capsule has been shown to be an important virulence factor of this organism by virtue of its ability to resist phagocytosis (Wessels *et al.*, 1991; Wessels *et al.*, 1994). The hyaluronidases from *S. pyogenes* bacteriophage H4489A (*hylP* and *hylP2*) are thought to be required to degrade the HA capsule that surrounds the bacterial host cell, thereby facilitating attachment of phage to the cell wall (Hynes *et al.*, 1995; Hynes and Walton, 2000). *HylP* contains a series of ten Gly-X-Y amino acid triplets, closely resembling the characteristic repeating sequences found in collagen (Stern and Stern, 1992; Hynes *et al.*, 1995). A possible role in human disease has been suggested by Hynes *et al.* (1995). This collagen-like repeat found in the bacteriophage hyaluronidase could induce antibodies, which may cross-react with

tissue collagen and result in the polyarthritis often associated with rheumatic fever.

Production by *S. pyogenes* of both a HA capsule and hyaluronidase enzymatic activity capable of destroying the capsule is an interesting, yet unexplained, phenomenon. It is possible to speculate that the three hyaluronidases (HylP1, HylP2 and HylP3) expressed by *S. pyogenes* SF370 confer virulence, but as they are known to be phage-associated have, through evolution of the bacterium, been commandeered into the bacterial genome from bacteriophages as a strategy of invasion. As with *hylP* from bacteriophage H4489A, HylP2 and HylP3 from *S. pyogenes* SF370 contain a collagen-like repeat within the proteins. This ten-triplet repeat is the minimum length required for maintaining thermodynamic stability of a triple-helical collagen molecule (Csóka *et al.*, 1997). As yet, the function of these sequences in HylP2 and HylP3 is unknown. It is possible that this collagen domain may be able to interact with intact collagen molecules found in host tissues. If hyaluronidases HylP2 and HylP3 are secreted extracellulary by the bacterium, it seems possible that these enzymes could function as a virulence factor enhancing initial attachment of the bacteria to the host collagenous tissues and then facilitating subsequent tissue invasion, as speculated for bacterial hyaluronidases (Csóka *et al.*, 1997). Production of hyaluronidase by group A streptococci has long been suggested to aid the organism in its spread through the connective tissue (McClean, 1941) hence, hyaluronidase has been designated as one of the spreading factors of microbial origin (Hynes *et al.*, 1995).

6.7 Summary

HA, a major polysaccharide component of human connective tissue, is targeted by virulence factors of many human bacterial pathogens, such as *S. pyogenes*. The development of inhibitors to key virulence factors expressed by the invading organisms, thus preventing their establishment is essential. A structural and mechanistic understanding of these enzymes will ultimately facilitate this goal.

To summarise, HA-degrading enzymes have been shown to be present in a wide range of organisms from GH families 56, 69 and 84. This present study shows that ORFs HylP1, HylP2 and HylP3 from *S. pyogenes* SF370 do indeed code for

active hyaluronidases. Characterisation of the N-terminally tagged HylP1, HylP2 and HylP3 revealed an optimum pH of 6.5, 6.0, and 5.5, respectively, highest amount of activity was observed at 37 °C and none of the three enzymes were very thermostable.

Progress towards a fundamental comprehension of protein/carbohydrate interactions of GH69 enzymes, which facilitate glycosidic bond cleavage, is currently hampered by a lack of 3D structural information. Recent structural studies on HylP1 from *S. pyogenes* SF370 has led to the elucidation of the 3D structure. The structure of HylP1 resembles a phage tail protein and although comparison with the bacteriophage T4 structure shows they are different, they appear to share some common features such as this phage tail structure. Also from comparisons with bacteriophage T4 it is possible to speculate that the phage-like tail from HylP1 adheres to the cell membrane of *S. pyogenes* to enable the hyaluronidase to facilitate the degradation of the HA capsule surrounding the bacterium.

Of interest is the presence of a collagenous domain in HylP2 and HylP3, which is of sufficient length to potentially provide stability for a triple helical structure. This collagen-like domain is also present in *S. pyogenes* bacteriophage H4489A, so a greater understanding as to why this region is present in various organisms is needed. Both primary sequence analysis and structural data is providing evidence of divergent evolution between *S. pyogenes* SF370 hyaluronidases and bacteriophage-associated hyaluronidases.

6.8 Future studies

Numerous opportunities for future studies exist in the areas of research covered in this thesis. To identify the location of the active site in HylP1, a mutagenesis strategy would ultimately need to be employed. Armed with a detailed biochemical analysis together with sequence homology data for GH69, site-directed mutants of HylP1, HylP2 and HylP3 could be constructed and biochemically characterised, providing valuable data on the relative importance of highly conserved amino acids. Using the recently solved 3D structure of HylP1 would then allow the location of mutated amino acids within the active site to be

identified. Finally, crystallisation of an inactive mutant in complex with a HA oligomer would allow the location of conserved amino acids in relation to the substrate, hence potentially elucidating the roles of conserved amino acids.

Although HylP1, HylP2 and HylP3 have been shown to use an endo mode of action, there is still a great amount to be learned about the mechanism for these family 69 enzymes. HylP1, HylP2 and HylP3 could be further characterised using high performance anion exchange chromatography (HPAEC) in terms of its depolymerisation pattern exhibited against HA. This will provide important insights into the binding mechanism, whether it is retaining or inverting, for a family for which there is currently no information.

Overall aim would be to study the structure/function relationship of hyaluronidases produced by *S. pyogenes* SF370 in detail in order to understand what renders closely related enzymes active at different pH or under varying environmental conditions.

7 References

Abou-Hachem, M., Nordberg Karlsson, E., Bartonek-Roxå, E., Raghothama, S., Simpson, P.J., Gilbert, H.J., Williamson, M.P., and Holst, O. (2000). Carbohydrate-binding modules from a thermostable *Rhodothermus marinus* xylanase: cloning, expression and binding studies. *Biochemical Journal.* **345**, 53-60.

Abou-Hachem, M., Nordberg Karlsson, Simpson, P.J., Linse, S., Sellers, P., Williamson, M.P., Jamieson, S.J., Gilbert, H.J., Bolam, D.N., and Holst, O. (2002). Calcium binding and thermostability of carbohydrate binding module CBM4-2 of Xyn10A from *Rhodothermus marinus*. *Biochemistry.* **41**, 5720-5729.

Albersheim, P., An, J., Freshour, G., Fuller, M.S., Guillen, R., Ham, K.S., Hahn, M.G., Huang, J., O'Neill, M., Whitcombe, A., Williams, M.V., York, W.S., and Darvill, A. (1994). Structure and function studies of plant cell wall polysaccharides. *Biochemical Society Transactions.* **22**, 374-378.

Alberti, S., Ashbaugh, C.D., and Wessels, M.R. (1998). Structure of the has operon promoter and regulation of hyaluronic acid capsule expression in group A Streptococcus. *Molecular Microbiology.* **28** (2), 343-353.

Alberts, B., Bray, D., Lewis, J., Raff, M., Roberts, K., and Watson, J.D. (1989). Molecular Biology of the Cell. Third Edition. Garland Publishing, Inc.

Aleshin, A., Golubev, A., Firsov, L.M., and Honzatko, R.B. (1992). Crystal structure of glucoamylase from *Aspergillus awamori* var. X100 to 2.2 Å resolution. *The Journal of Biological Chemistry.* **267**, 19291-19298.

Alkrad, J. A., Mrestani, Y., Stroehl, D., Wartewig, S., and Neubert, R. (2003). Characterization of enzymatically digested hyaluronic acid using NMR, Raman, IR, and UV-Vis spectrocopies. *Journal of Pharmaceutical and Biomedical Analysis.* **31**, 545-550.

Alouf, J.E. (1980). Streptococcal toxins (streptolysin O, streptolysin S, erythrogenic toxin). *Pharmacology Therapy.* **11** (3), 661-717.

Alouf, J.E., and Loridan, C. (1988). Production, purification, and assay of streptolysin S. *Methods in Enzymology.* **165**, 59-64.

Arai, T., Araki, R., Tanaka, A., Karita, S., Kimura, T., Sakka, K., and Ohmiya, K. (2003). Characterization of a cellulase containing a family 30 carbohydrate-binding module (CBM) derived from *Clostridium thermocellum* CelJ: importance of the CBM to cellulose hydrolysis. *Journal of Bacteriology.* **185** (2), 504-512.

Arming, S., Strobl, B., Wechselberger, C., and Kreil, G. (1997). *In vitro* mutagenesis of PH-20 hyaluronidase from human sperm. *European Journal of Biochemistry.* **247**, 810-814.

Ashbaugh, C.D., Alberti, S., and Wessels, M.R. (1998). Molecular analysis of the capsule gene region of group A Streptococcus: the hasAB genes are sufficient for capsule expression. *Journal of Bacteriology*. **180** (18), 4955-4959.

Atkins, E.D.T. (1992). Three-dimensional structure, interactions and properties of xylans. In: *Progress in Biotechnology, Vol. 7: Xylans and Xylanases*, 39-50. J. Visser, G. Beldman, M.A. Kuster-van Someren and A.G.J. Voragen (Eds.). Elsevier, Amsterdam.

Bae, H-J., Turcotte, G., Chamberland, H., Karita, S., Vézina, L-P. (2003). A comparative study between an endoglucanase IV and its fused protein complex Cel5-CBM6. *FEMS Microbiology Letters*. **227**, 175-181.

Baker, J.R., Yu, H., Morrison, K., Averett, W.F., and Pritchard. D.G. (1997). Specificity of the hyaluronate lyase of group-B streptococcus toward unsulphated regions of chondroitin sulphate. *Biochemical Journal*. **327** (1), 65-71.

Baker, J.R., Dong, S., and Pritchard, D.G. (2002). The hyaluronan lyase of *Streptococcus pyogenes* bacteriophage H4489A. *Biochemical Journal*. **365**, 317-322.

Banks, D.J., Lei, B., and Musser, J.M. (2003). Prophage induction and expression of prophage-encoded virulence factors in group A *Streptococcus* serotype M3 strain MGAS315. *Infection and Immunity*. **71** (12), 7079-7086.

Banumathi, S., Dauter, M., and Dauter, Z. (2003). Phasing at high resolution using Ta_6Br_{12} cluster. *Acta Crystallographica, Section D, Biological Crystallography*. **D59**, 492-498.

Bartlett, M.R., Cowden, W.B., and Parish, C.R. (1995). Differential effects of the anti-inflammatory compounds heparin, mannose-6-phosphate, and castanospermine on degradation of the vascular basement membrane by leukocytes, endothelial cells, and platelets. *Journal of Leukocyte Biology*. **57** (2), 207-213.

Bartolazzi, A., Peach, R., Aruffo, A., and Stamenkovic, I. (1994). Interaction between CD44 and hyaluronate is directly implicated in the regulation of tumor-development. *Journal of Experimental Medicine*. **180**, 53-66.

Béguin, P., and Alzari, P.M. (1998). The cellulosome of *Clostridium thermocellum*. *Biochemical Society Transactions*. **26**, 178-185.

Béguin, P., and Aubert, J. P. (1994). The biological degradation of cellulose. *FEMS Microbiology Review*. **13**, 25-58.

Beguin, P., and Lemaire, M. (1996). The cellulosome: an exocellular, multiprotein complex specialized in cellulose degradation. *Critical Reviews in Biochemistry and Molecular Biology*. **31** (3), 201-236.

Berisio, R., Granata, V., Vitagliano, L., and Zagari, A. (2004). Characterization of collagen-like heterotrimers: implications for triple-helix stability. *Biopolymers*. **73** (6), 682-688.

Bernheimer, A.W., Lim, K.S., Remsen, C.C., Antanavage, J., and Watson, S.W. (1972). Factors affecting interaction of staphylococcal alpha toxin with membranes. *Infection and Immunity*. **6** (4), 636-642.

Berry, A.M., Lock, R.A., Thomas, S.M., Rajan, D.P., Hansman, D., and Paton, J.C. (1994). Cloning and nucleotide sequence of the *Streptococcus pneumoniae* hyaluronidase gene and purification of the enzyme from recombinant *Escherichia coli*. *Infection and Immunity*. **62** (3), 1101-1108.

Betley, M.J., Borst, D.W., and Regassa, L.B. (1992). Staphylococcal enterotoxins, toxic shock syndrome toxin and streptococcal pyrogenic exotoxins: a comparative study of their molecular biology. *Chemical Immunology*. **55**, 1-35.

Biely, P., Côté, G.L., Kremnický, L., and Greene, R.V. (1999). Differences in catalytic properties of acetylxylan esterases and non-hemicellulolytic acetylesterases. In: *Recent advances in carbohydrate bioengineering,* pp 73-81. H. J. Gilbert, G. J. Davies, B. Henrissat and B. Svensson (Eds.). The Royal Society of Chemistry, Cambridge.

Birsan, C., Johnson, P., Joshi, M., MacLeod, A., McIntosh, L., Monem, V., Nitz, M., Rose, D.R., Tull, D., Wakarchuck, W.W., Wang, Q., Warren, R.A.J., White, A., and Withers, S.G. (1998). Mechanisms of cellulases and xylanases. *Biochemical Society Transactions*. **26**, 156-160.

Bisno, A.L. (1979). Alternate complment pathway activation by group A streptococci: role of M-protein. *Infection and Immunity*. **26** (3), 1172-1176.

Bisno, A.L., Brito, M.O., and Collins, C.M. (2003). Molecular basis of group A streptococcal virulence. *The Lancet: Infectious Diseases*. **3**, 191-200.

Bissell, M.J., and Nelson, W.J. (1999). Cell-to-cell contact and extracellular matrix integration of form and function: the central role of adhesion molecules. *Current opinion in Cell Biology*. **11**, 537-539.

Blaak, H., and Schrempf, H. (1995). Binding and substrate specificities of a *Streptomyces olivaceoviridis* chitinase in comparison with its proteolytically processed form. *European Journal of Biochemistry*. **229**, 132-139.

Black, G.W., Hazlewood, G. P., Millward-Saddler, S.J., Laurie, J.I., and Gilbert, H. J. (1995). A modular xylanase containing a novel non-catalytic xylan-specific binding domain. *Biochemical Journal*. **307**, 191-195.

Black, G.W., Rixon, J.E., Clarke, J.H., Hazlewood, G.P., Theodorou, M.K., Morris, P., and Gilbert, H.J. (1996). Evidence that linker sequences and cellulose binding domains enhance the activity of hemicellulases against complex substrates. *Biochemical Journal*. **319**, 515-520.

Bolam, D. N., Ciruela, A., McQueen-Mason, S., Simpson, P., Williamson, M. P., Rixon, J. E., Boraston, A., Hazlewood, G. P., and Gilbert, H. J. (1998). *Pseudomonas* cellulose-binding domains mediate their effects by increasing enzyme substrate proximity. *Biochemical Journal.* **331**, 775-781.

Bolam, D. N., Xie, H., White, P., Simpson, P.J., Hancock, S.M., Williamson, M.P., and Gilbert, H.J. (2001). Evidence for synergy between family 2b carbohydrate binding modules in *Cellulomonas fimi* xylanase 11A. *Biochemistry.* **40**, 2468-2477.

Boraston, A.B., McLean, B.W., Kormos, J.M., Alam, M., Gilkes, N.R., Haynes, C.A., Tomme, P., Kilburn, D.G., and Warren, R.A.J. (1999). Carbohydrate-binding modules: diversity of structure and function. In: *Recent Advances in Carbohydrate Bioengineering*, pp 202-211. H.J. Gilbert, G.J. Davies, B. Henrissat, and B. Svensson (Eds.). The Royal Society of Chemistry, Cambridge.

Boraston, A.B., Chiu, P., Warren, R.A.J., and Kilburn, D.G. (2000a). Specificity and affinity of substrate binding by a family 17 carbohydrate-binding module from *Clostridium cellulovorans* cellulase 5A. *Biochemistry.* **39**, 11129-11136.

Boraston, A.B., Tomme, P., Amandoran, E.A., and Kilburn, D.G. (2000b). A novel mechanism of xylan binding by a lectin-like module from *Streptomyces lividans* xylanase 10A. *Biochemical Journal.* **350**, 933-941.

Boraston, A.B., Warren, R.A., and Kilburn DG. (2001). Beta-1,3-Glucan binding by a thermostable carbohydrate-binding module from *Thermotoga maritima*. *Biochemistry.* **40** (48), 14679-14685.

Boraston, A.B., Creagh, A.L., Alam, M.M., Kormos, J.M., Tomme, P., Haynes, C.A., Warren, R.A., and Kilburn, D.G. (2001). Binding specificity and thermodynamics of a family 9 carbohydrate-binding module from *Thermotoga maritima* xylanase 10A. *Biochemistry.* **40** (21), 6240-6247.

Boraston, A.B., Nurizzo, D., Notenboom, V., Ducros, V., Rose, D.R., Kilburn, D.G., and Davies, G.J. (2002a). Differential oligosaccharide recognition by evolutionarily-related beta-1,4 and beta-1,3 glucan-binding modules. *Journal of Molecular Biology.* **319** (5), 1143-1156.

Boraston, A.B., Ghaffari, M., Warren, R.A., and Kilburn, D.G. (2002b). Identification and glucan-binding properties of a new carbohydrate-binding module family. *Biochemical Journal.* **361** (Pt 1), 35-40.

Boraston, A.B., Notenboom, V., Warren, R.A.J., Kilburn, D.G., Rose, D.R., and Davies, G.J. (2003). Structure and ligand binding of carbohydrate-binding module CsCBM6-3 reveals similarities with fucose-specific lectins and "galactose-binding" domains. *Journal of Molecular Biology.* **327** (3), 659-69.

Bradford, M. M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principal of protein-dye binding. *Analytical Biochemistry*. **72**, 248-254.

Bray, M.R., Johnson, P.E., Gilkes, N.R., McIntosh, L.P., Kilburn, D.G., and Warren, R.A. (1996). Probing the role of tryptophan residues in a cellulose-binding domain by chemical modification. *Protein Science*. **5** (11), 2311-2318.

Brect, M., Mayer, U., Schlosser, E., and Prehm, P. (1986). Increased hyaluronate synthesis is required for fibroblast detachment and mitosis. *Biochemical Journal*. **239**, 445-450.

Brett, C. T. and Waldron, K. (1996). Physiology and biochemistry of plant cell walls. Topics in plant functional biology: 1. M. Black and B. Charlwood (Eds.). Second Edition. Chapman and Hall Publishers, London.

Brodsky, B., and Ramshaw, J.A. (1997). The collagen triple-helix structure. *Matrix Biology*. **15** (8-9), 545-554.

Brown, I. E., Mallen, M. H., Charnock, S. J., Davies, G. J. and Black, G. W. (2001). Pectate lyase 10A from *Pseudomonas cellulosa* is a modular enzyme containing a family 2a carbohydrate-binding module. *Biochemical Journal*. **355**, 155-165.

Brun, E., Johnson, P.E., Creagh, A.L., Tomme, P., Webster, P., Haynes, C.A., McIntosh, L.P. (2000). Structure and binding specificity of the second N-terminal cellulose-binding domain from *Cellulomonas fimi* endoglucanase C. *Biochemistry*. **39** (10), 2445-2458.

Buckwalter, J.A., and Manken, H.J. (1998). Articular cartilage repair and transplantation. *Arthritis and Rheumatism*. **41** (8), 1331-1342.

Buckwalter, J.A., and Manken, H.J. (1998). Articular cartilage: tissue design and chondrocyte matrix interactions. *Instructional Course Lectures*. **47**, 477-486.

Bullock, W.O., Fernandez, J.M., and Short, J.M. (1987). XL1-Blue – a high efficiency plasmid transforming *recA Escherichia coli* strain with β-galactosidase selection. *Biotechniques*. **5**, 376.

Calabro, A., Midura, R., Wang, A., West, L., Plaas, A., and Hascall, V.C. (2001). Flurophore-assisted carbohydrate electrophoresis (FACE) of glycosaminoglycans. *Osteoarthritis and Cartilage*. **9**, Supplement A, S16-S22.

Campbell, J.A., Davies, G.J., Bulone, V., and Henrissat, B. (1997). A classification of nucleotide-diphospho-sugar glycosyltransferases based on amino acid sequence similarities. *Biochemical Journal*. **326**, 929-942.

Canard, B., Garnier, T., Saint-Joanis, B., and Cole, S.T. (1994). Molecular genetic analysis of the *nagH* gene encoding a hyaluronidase of *Clostridium perfringens*. *Molecular and General Genetics*. **243**, 215-224.

Caparon, M.G., Stephens, D.S., Olsen, A., and Scott, J.R. (1991). Role of M protein in adherence of group A streptococci. *Infection and Immunity*. **59**, 1811-1817.

Carpita, N. C. and Gilbeaut, D. M. (1993). Structural models of primary cell walls in flowering plant: consistency of molecular structure with the physical properties of the walls during growth. *The Plant Journal*. **3** (1), 1-30.

Carpita, N. (1997). Structure and biosynthesis of plant cell walls. In: Plant Metabolism. pp 124-160. D.T., Dennis; D.H., Turpin; D.D., Lefebvre, and D.B., Layzell (Eds.). Second Edition. Longman.

Carpita, N. C., and Vergara, C. (1998). A recipe for cellulose. *Science*. **279**, 672-673.

Carvalho, A.L., Dias, F.M.V., Prates, J.A.M., Nagy, T., Gilbert, H.J., Davies, G.J., Ferreira, L.M.A., Romão, M.J., and Fontes, C.M.G.A. (2003). Cellulosome assembly revealed by the crystal structure of the cohesin-dockerin complex. *Proceedings of the National Academy of Sciences, USA*. **100** (24), 13809-13814.

Cashman, D.C., Laryea, J.U., and Weissmann, B. (1969). The hyaluronidase of rat skin. *Archives of Biochemistry and Biophysics*. **135** (1), 387-395.

CCP4 (Collaborative Computational Project Number 4; 1994). The CCP4 suite: programs for protein crystallography. *Acta Crystallographica Section D Biological Crystallography*. **50**, 760-763.

Chain, E.A., and Duthie, E.S. (1939). A mucolytic enzyme in testes extract. *Nature*. **144**, 977-978

Chain, E., and Duthie, E.S. (1940). Identity of hyaluronidase and spreading factor. *Journal of Experimental Pathology*. **Xxi**, 324-338.

Chan, V.C., Ramshaw, J.A.M., Kirkpatrick, A., Beck, K., and Brodsky, B. (1997). Positional preferences of ionisable residues in Gly-*X*-*Y* triplets of the collagen triple-helix. *The Journal of Biological Chemistry*. **272** (50), 31441-31446.

Charnock, S. J., Lakey, J.H., Virden, R., Hughes, N., Sinnott, M.L., Hazlewood, G.P., Pickersgill, R., and Gilbert, H.J. (1997). Key residues in subsite F play a critical role in the activity of *Pseudomonas fluorescens* subspecies *cellulose* xylanase A against xylooligosaccharides but not against highly polymeric substrates such as xylan. *The Journal of Biological Chemistry*. **272** (5), 2942-2951.

Charnock, S. J., Spurway, T. D., Xie, H., Beylot, M-H., Virden, R., Warren, R. A. J., Hazelwood, G. P., and Gilbert, H. J. (1998). The topology of the substrate binding clefts of glycosyl hydrolase family 10 xylanases are not conserved. *The Journal of Biological Chemistry*. **273** (48), 32187-32199.

Charnock, S. J., and Davies, G.J. (1999). Structure of the nucleotide-diphospho-sugar transferase, SpsA from *Bacillus subtilis*, in native and complexed forms. *Biochemistry*. **38** (20), 6380-6385.

Charnock, S. J., Bolam, D.N., Turkenburg, J.P., Gilbert, H.J., Ferreira, L.M.A, Davies, G.J., and Fontes, C.M.G.A. (2000). The X6 "thermostabilizing" domains of xylanases are carbohydrate-binding modules: structure and biochemistry of the *Clostridium thermocellum* X6b domain. *Biochemistry*. **39**, 5013-5021.

Charnock, S. J., Brown, I. E., Turkenburg, J. P., Black, G. W. & Davies, G. J. (2002a) Convergent evolution sheds light on the *anti* β-elimination mechanism common to family 1 and 10 polysaccharide lyases. *Proceedings of the National Academy of Sciences, USA*. **99**, (19), 12067-12072.

Charnock, S.J., Bolam, D.N., Nurizzo, D., Szabó, L., McKie, V.A., Gilbert, H.J., and Davies, G.J. (2002b). Promiscuity in ligand-binding: The three-dimensional structure of a *Piromyces* carbohydrate-binding module, CBM29-2, in complex with cello- and mannohexaose. *Proceedings of the National Academy of Sciences, USA*. **99** (22), 14077-14082.

Cherr, G.N., Yudin, A.I., and Overstreet, J.W. (2001). The dual functions of GPI-anchored PH-20: hyaluronidase and intracellular signalling. *Matrix Biology*. **20**, 515-525.

Clarke, J.H., Laurie, J.I., Gilbert, H.J., and Hazlewood, G.P. (1991). Multiple xylanases of *Cellulomonas fimi* are encoded by distinct genes. *FEMS Microbiology Letters*. **83**, 305-310.

Clarke, M.R., Davidson, K., Gilbert, H.J., Fontes, C.M.G.A., and Hazlewood, G.P. (1996). A modular xylanase from mesophilic *Cellulomonas fimi* contains the same cellulose-binding and thermostabilising domains as xylanases from thermophilic bacteria. *FEMS Microbiology*. **139**, 27-35.

Coggins, J.R. (1991). Deletions, fusions and domain rearrangements. *Current Opinion in Biotechnology*. **2**, 576-581.

Coleman, P.J. (2002). Evidence for a role of hyaluronan in the spacing of fibrils within collagen bundles in rabbit synovium. *Biochimica et Biophysica Acta*. **1571**, 173-182.

Cosgrove, D.J. (1997). Assembly and enlargement of the primary cell wall in plants. *Annual Review of Cell and Develpomental Biology*. **13**, 171-201.

Coughlan, M.P. (1985). The properties of fungal and bacterial cellulases with comment on their production and application. In: *Biotechnology and genetic engineering reviews*, Vol. 3, 39-109. G.E. Russell (Ed.). Intercept.

Coutinho, P. M., and Henrissat, B. (1999). Carbohydrate-active enzymes server at URL: http://afmb.cnrs-mrs.fr/~cazy/CAZY/index.html

Coutinho, P. M., and Henrissat, B. (1999). Carbohydrate-Active Enzymes: an integrated database approach. In: *Recent advances in carbohydrate bioengineering,* pp 3-12. H. J. Gilbert, G. J. Davies, B. Henrissat and B. Svensson (Eds.). The Royal Society of Chemistry, Cambridge.

Coutinho, P. M., Deleury, E., Davies, G.J., and Henrissat, B. (2003). An evolving hierarchical family classification for glycosyltransferases. *Journal of Molecular Biology*. **328**, 307-317.

Cramer, J.A., and Bailey, L.C. (1991). A reversed-phase ion-pair high-performance liquid chromatography method for bovine testicular hyaluronidase digests using postcolumn derivatization with 2-cyanoacetamide and ultra-violet detection. *Analytical Biochemistry*. **196**, 183-191.

Cramer, J.A., Bailey, L.C., Bailey, C.A., and Miller, R.T. (1994). Kinetic and mechanistic studies with bovine testicular hyaluronidase. *Biochimica et Biophysica Acta*. **1200**, 315-321.

Crater, D.L., and van de Rijn, I. (1995). Hyaluronic acid synthesis operon (has) expression in group A streptococci. *The Journal of Biological Chemistry*. **270** (31), 18452-18458.

Creagh, A.L., Ong, E., Jervis, E., Kilburn, D.G., and Haynes, C.A. (1996). Binding of the cellulose-binding domain of exoglucanase Cex from *Cellulomonas fimi* to insoluble microcrystalline cellulose is entropically driven. *Proceedings of the National Academy of Sciences, USA*. **93** (22), 12229-12234.

Csóka, A.B., Frost, G.I., Wong, T., Stern, R., and Csóka, T.B. (1997). Purification and microsequencing of hyaluronidase isozymes from human urine. *FEBS Letters*. **417**, 307-310.

Csóka, T.B., Frost, G.I., Heng, H.H., Scherer, S.W., Mohapatra, G., and Stern, R. (1998). The hyaluronidase gene HYAL1 maps to chromosome 3p21.2-p21.3 in human and 9F1-F2 in mouse, a conserved candidate tumor suppressor locus. *Genomics*. **48** (1), 63-70.

Csóka, A.B., Frost, G.I., and Stern, R. (2001). The six hyaluronidase-like genes in the human and mouse genomes. *Matrix Biology*. **20**, 499-508.

Culty, M., Nguyen, H.A., and Underhill, C.B. (1992). The hyaluronan receptor (CD44) participates in the uptake and degradation of hyaluronan. *Journal of Cellular Biology*. **116** (4), 1055-1062.

Cunningham, M.W. (2000). Pathogenesis of group A Streptococcal infections. *Clinical Microbiogy Reviews*. **13** (3), 470-511.

Czjzek, M., Bolam, D.N., Mosbah, A., Allouch, J., Fontes, C.M.G.A., Ferreira, L.M.A., Bornet, O., Zamboni, V., Darbon, H., Smith, N.L., Black, G.W., Henrissat, B., and Gilbert, H.J. (2001). The location of the ligand-binding site of

carbohydrate-binding modules that have evolved from a common sequence is not conserved. *The Journal of Biological Chemistry.* **276** (51), 48580-48587.

Dale, J.B., Washburn, R.G., Marques, M.B., and Wessels, M.R. (1996). Hyaluronate capsule and surface M protein in resistance to opsonization of group A streptococci. *Infection and Immunity.* **64** (5), 1495-1501.

Dale, J.B., Chiang, E.Y., Hasty, D.L., and Courtney, H.S. (2002). Antibodies against a synthetic peptide of SagA neutralize the cytolytic activity of streptolysin S from group A streptococci. *Infection and Immunity.* **70**: 2166-2170.

Davies, G. J. (1998). Structural studies on cellulases. *Biochemical Society Transactions.* **26**, 167-173.

Davies, G. J. and Henrissat, B. (1995). Structures and mechanisms of glycosyl hydrolases. *Structure.* **3** (9), 853-859.

Davies, G.J., and Charnock, S.J. (1999). Structure of a nucleotide-diphospho-sugar transferase: implications for the synthesis of polysaccharides. In: *Recent advances in carbohydrate bioengineering,* pp 132-143. H. J. Gilbert, G. J. Davies, B. Henrissat and B. Svensson (Eds.). The Royal Society of Chemistry, Cambridge.

Davies, G. J. and Henrissat, B. (2002). Structural enzymology of carbohydrate-active enzymes: implications for the post-genomic era. *Biochemical Society Transactions.* **30** (2), 291-297.

Davies, G. J., Wilson, K. S. and Henrissat, B. (1997). Nomenclature for sugar-binding subsites in glycosyl hydrolases. *Biochemical Journal.* **321**, 557-559.

Day, A.J. (1999). The structure and regulation of hyaluronan-binding proteins. *Biochemical Society Transactions.* **27**, 115-121.

DeAngelis, P.L. (1999). Molecular directionality of polysaccharide polymerisation by the *Pasturella multocida* hyaluronan synthase. *The Journal of Biological Chemistry.* **274** (37), 26557-26562.

Dempsey, L.A., Plummer, T.B., Coombes, S.L., and Platt, J.L. (2000). Heparanase expression in invasive trophoblasts and acute vascular damage. *Glycobiology.* **10** (5), 467-475.

Devillard, E., Bera-Maillet, C., Flint, H.J., Scott, K.P., Newbold, C.J., Wallace, R.J., Jouany, J.P., and Forano, E. (2003). Characterization of XYN10B, a modular xylanase from the ruminal protozoan *Polyplastron multivesiculatum*, with a family 22 carbohydrate-binding module that binds to cellulose. *Biochemical Journal.* **373** (Pt 2), 495-503.

Dick, G.F., and Dick, G.H. (1924). Landmark article Jan 26, 1924: The etiology of scarlet fever. *Journal of the American Medical Association.* **250** (22), 3096.

Din, N., Gilkes, N.R., Tekant, B., Miller, R.C., Warren, R.A.J., and Kilburn, D.G. (1991). Non-hydrolytic disruption of cellulose fibres by the binding domain of a bacterial cellulase. *Bio/Technology*. **9**, 1096-1099.

Din, N., Forsythe, I.J., Burtnick, L.J., Gilkes, N.R., Miller, R.C., Warren, R.A.J., and Kilburn, D.G. (1994). The cellulose binding domain of endoglucanase A (CenA) from *Cellulomonas fimi*: evidence for the involvement of tryptophan residues in binding. *Molecular Microbiology*. **11**, 747-755.

Din, N., Coutinho, J.B., Gilkes, N.R., Jervis, E., Kilburn, R.C., Miller, R.C., Jr., Ong, E., Tomme, P., anmd Warren, R.A.J. (1995). Interactions of cellulases from *Cellulomonas fimi* with cellulose. In: *Progress in Biotechnology*, Vol 10: *Carbohydrate Bioengineering*, 261-270. S.B. Petersen, B. Svensson, and S. Pedersen (Eds.). Elsevier, Amsterdam.

Divne, C., Stahlberg, J., Reinikainen, T., Ruohenen, L., Petterson, G., Knowles, J.K.C., Teeri, T.T., and Jones, T.A. (1994). The three dimensional crystal structure of the catalytic core of cellobiohydrolase I from *Trichoderma reesei*. *Science*. **265**, 524-528.

Dupont, C., Roberge, M., Shareck, F., Morosoli, R., and Kluepfel, D. (1998). Substrate-binding domains of glycanases from *Streptomyces lividans*: characterization of a new family of xylan-binding domains. *Biochemical Journal*. **330**, 41-45.

Emami, K., Nagy, T., Fontes, C. M. G. A., Ferreira, L. M. A. and Glibert, H. J. (2002) Evidence for temporal regulation of the two *Pseudomonas cellulosa* xylanases belonging to glycoside hydrolase family 11. *Journal Bacteriology*. **184** (15), 4124-4133.

Ernst, S., Langer, R., Cooney, C.L., and Sasisekharan, R. (1995). Enzymatic degradation of glycosaminoglycans. *Critical Reviews in Biochemistry and Molecular Biology*. **30** (5), 387-444.

Esko, J.D., and Lindahl, U. (2001). Molecular diversity of heparan sulfate. *The Journal of Clinical Investigation*. **108** (2), 169-173.

Fath, M., VanderNoot, V., Kilpeläinen, I., Kinnunen, T., Rauvala, H., and Linhardt, R.J. (1999). Interaction of soluble and surface-bound heparin binding growth-associated molecule with heparin. *FEBS Letters*. **454**, 105-108.

Fernandes, A.C., Fontes, C.M.G.A., Gilbert, H.J., Hazlewood, G.P., Fernandes, T.H., and Ferreira, L.M.A. (1999). Homologous xylanases from *Clostridium thermocellum*: evidence for bi-functional activity, synergism between xylanase catalytic modules and the presence of xylan-binding domains in enzyme complexes. *Biochemical Journal*. **342**, 105-110.

Ferreira, L.M.A., Wood, T.M., Williamson, G., Faulds, C., Hazlewood, G.P., Black, G.W., and Gilbert, H.J. (1993). A modular esterase from *Pseudomonas*

fluroescens subsp. *cellulosa* contains a non-catalytic cellulose-binding domain. *Biochemical Journal.* **294**, 349-355.

Ferretti, J.J., McShan, W.M., Ajdic, D., Savic, D.J., Savic, G., Lyon, K., Primeaux, C., Sezate, S., Suvorov, A.N., Kenton, S., Lai, H.S., Lin, S.P., Qian, Y., Jia, H.G., Najar, F.Z., Ren, Q., Zhu, H., Song, L., White, J., Yuan, X., Clifton, S.W., Roe, B.A., and McLaughlin, R. (2001). Complete genome sequence of an M1 strain of *Streptococcus pyogenes. Proceedings of the National Academy of Sciences, USA.* **98** (8), 4658-4663.

Féthière, J., Eggimann, B., and Cygler, M. (1999). Crystal structure of chondroitin AC lyase, a representative of a family of glycosaminoglycan degrading enzymes. *Journal of Molecular Biology.* **288** (4), 635-647.

Fierobe, H.P., Bayer, E.A., Tardif, C., Czjzek, M., Mechaly, A., Belaich, A., Lamed, R., Shoham, Y., and Belaich, J.P. (2002). Degradation of cellulose substrates by cellulosome chimeras. Substrate targeting versus proximity of enzyme components.
The Journal of Biological Chemistry. **277** (51), 49621-49630.

Fiszer-Szafarz, B., Litynska, A., and Zou, L. (2000). Human hyaluronidases: electrophoretic multiple forms in somatic tissues and body fluids. Evidence for conserved hyaluronidase potential *N*-glycosylation sites in different mammalian species. *Journal of Biochemical and Biophysical Methods.* **45**, 103-116.

Fitzgerald, T/J., and Gannon, E.M. (1983). Further evidence for hyaluronidase activity of *Treponema pallidum. Canadian Journal of Microbiology.* **29**, 1507-1513.

Foley, M.J., and Wood, W.B. (1959). Studies of the pathogenicity of group A streptococci. II. The anti-phagocytic effects of the M protein and its capsular gel. *Journal of Experimental Medicine.* **110**, 617-628.

Fontes, C.M.G.A, Hazlewood, G.P., Morag, E., Hall, J., Hirst, B.H., and Gilbert H.J. (1995). Evidence for a general role for non-catalytic thermostabilizing domains in xylanases from thermophilic bacteria. *Biochemical. Journal.* **307**, 151-158.

Fraser, J.D. (1992). Superantigen data. *Nature.* **360** (6403), 423.

Fraser, J.R.E., Laurent, T.C., and Laurent, U.B.G. (1997). Hyaluronan: its nature, distribution, functions and turnover. *Journal of Internal Medicine.* **242**, 27-33.

Fraser, J.R.E., and Laurent, T.C. (1997). Hyaluronan. Minisymposium. *Journal of Internal Medicine.* **242**, 25-26.

Freelove, A.C., Bolam, D.N., White, P., Hazlewood, G.P., and Gilbert, H.J. (2001). A novel carbohydrate-binding protein is a component of the plant cell wall-degrading complex of *Piromyces equi. Journal of Biological Chemistry.* **276** (46), 43010-43017.

Freeman, C., and Parish, C.R. (1998). Human platelet heparanase: purification, characterization and catalytic activity. *Biochemical Journal.* **330** (3), 1341-1350.

Frost, G.I., Csóka, A.B., Stern, R., and Csóka, T.B. (1996). The hyaluronidases: a chemical, biological and clinical overview. *Trends in Glycoscience and Glycotechnology.* **8** (44), 419-434.

Frost, G.I., and Stern, R. (1997). A microtiter-based assay for hyaluronidase activity not requiring specialized reagents. *Analytical Biochemistry.* **251**, 263-269.

Frost, G.I., Csóka, T.B., Wong, T., and Stern, R. (1997). Purification, cloning, and expression of human plasma hyaluronidase. *Biochemical and Biophysical Research Communications.* **236**, 10-15.

Fuglsang, C.C., Berka, R.M., Wahleithner, J.A., Kauppinen, S., Shuster, J.R., Rasmussen, G., Halkier, T., Dalboge, H., Henrissat, B. (2000). Biochemical analysis of recombinant fungal mutanases. A new family of alpha 1,3-glucanases with novel carbohydrate-binding domains. *Journal of Biological Chemistry.* **275** (3), 2009-2018.

Fujimoto, Z., Kaneko, S., Kuno, A., Kobayashi, H., Kusakabe, I., and Mizuno, H. (2002). Crystal structures of the sugar complexes of *Streptomyces olivaceoviridis* E-86 xylanase: sugar binding structure of the family 13 carbohydrate binding module. *Journal of Molecular Biology.* **316** (1), 65-78.

Fujimoto, Z., Kuno, A., Kaneko, S., Kobayashi, H., Kusakabe, I., and Mizuno, H. (2004). Crystal structures of decorated xylooligosaccharides bound to a family 10 xylanase from *Streptomyces olivaceoviridis* E-86. *Journal of Biological Chemistry.* **279** (10), 9606-9614.

Fujino, T., Béguin, P., and Aubert, J.P. (1993). Organization of a *Clostridium thermocellum* gene cluster encoding the cellulosomal scaffolding protein CipA and a protein possibly involved in attachment of the cellulosome to the cell surface. *Journal of Bacteriology.* **175**, 1891-1899.

Fujita, N., Yaegashi, N., Ide, Y., Sato, S., Nakamura, M., Ishiwata, I., and Yajima, A. (1994). Expression of CD44 in normal human versus tumor endometrial tissue: possible implications of reduced expression of CD44 in lymph-vascular space involvement of cancer cells. *Cancer Research.* **45**, 3711-3717.

Ghangas, G.S., Hu, Y.-J., and Wilson, D.B. (1989). Cloning of *Thermomonospora fusca* xylanase and its expression in *Escherichia coli* and *Streptomyces lividans*. *Journal of Bacteriology.* **171**, 2963-2969.

Gibeaut, D.M., and Carpita, N.C. (1994). Biosynthesis of plant cell wall polysaccharides. *The Federation of American Societies for Experimental Biology Journal.* **8** (12), 904-915.

Gilad, R., Rabinovich, L., Yaron, S., Bayer, E.A., Lamed, R., Gilbert, H.J., and Shoham, Y. (2003). CelI, a noncellulosomal family 9 enzyme from *Clostridium thermocellum*, is a processive endoglucanase that degrades crystalline cellulose. *Journal of Bacteriology*. **185** (2), 391-398.

Gilbert, H. J., Hall, J., Hazlewood, G. P., Ferreira, and L.M. (1990). The N-terminal region of an endoglucanase from *Pseudomonase fluorescens* subsp. *cellulosa* constitutes a cellulose-binding domain that is distinct from the catalytic centre. *Molecular Microbiology*. **4** (5), 759-767.

Gilbert, H. J. and Hazlewood, G. P. (1993). Bacterial cellulases and xylanases. *Journal of General Microbiology*. **139**, 187- 194.

Gilkes, N.R., Warren, R.A.J., Miller, R.C. Jr, and Kilburn, D.G. (1988). Precise excision of the cellulose binding domain from two *Cellulomonas fimi* cellulases by a homologous protease and the effect on catalysis. *The Journal of Biological Chemistry*. **263**, 10401-10407.

Gilkes, N.R., Henrissat, B., Kilburn, D.G., Miller, R.C. Jr, and Warren, R.A.J. (1991). Domains in microbial β-1,4-glycanases: sequence conservation, function and enzyme families. *Microbiology Review*. **55**, 303-315.

Gill, S.C., and von Hippel, P.H. (1989). Calculation of protein extinction coefficients from amino acid sequence data. *Analytical Biochemistry*. **182**, 319-326.

Gill, J., Rixon, J. E., Bolam, D. N., McQueen-Mason, S., Simpson, P. J., Williamson, M. P., Hazlewood, G. P. and Gilbert, H. J. (1999). The type II and X cellulose-binding domains of *Pseudomonas* xylanase A potentiate catalytic activity against complex substrates by a common mechanism. *Biochemical Journal*. **342**, 473-480.

Girish, K.S., Jagadeesha, D.K., Rajeev, K.B., and Kemparaju, K. (2002). Snake venom hyaluronidase: an evidence for isoforms and extracellular matrix degradation. *Molecular and Cellular Biochemistry*. **240**, 105-110.

Gmachl, M., and Kreil, G. (1993). Bee venom hyaluronidase is homologous to a membrane protein of mammalian sperm. *Proceedings of the National Academy of Sciences, USA*. **90**, 3569-3573.

Gmachl, M., Sagan, S., Ketter, S., and Kreil, G. (1993). The human sperm protein PH-20 has hyaluronidase activity. *FEBS Letters*. **336** (3), 545-548.

Godavarti, R., and Sasisekharan, R. (1996). A comparitive analysis of the primary sequences and characteristics of heparinases I, II, and III from *Flavobacterium heparinum*. *Biochemical and Biophysical Research Communications*. **229**, 770-777.

Godavarti, R., Davis, M., Venkataraman, G., Cooney, C., Langer, R., and Sasisekharan, R. (1996). Heparinase III from *Flavobacterium heparinum*: cloning

and recombinant expression in *Escherichia coli*. *Biochemical and Biophysical Research Communications*. **225**, 751-758.

Goldstein, M.A., Takagi, M., Hashida, S., Shoseyov, O., Doi, R.H., and Segel, I.H. (1993). Characterization of the cellulose-binding domain of the *Clostridium cellulovorans* cellulose-binding protein A. *Journal of Bacteriology*. **175** (18), 5762-5768.

Gorham, S.D., Olavesen, H., and Dodgson, K.S. (1975). Effect of ionic strength and pH on the properties of purified bovine testicular hyaluronidase. *Connective Tissue Research*. **3**, 17-25.

Gosalbes, M.J., Perez-Gonzalez, J.A., Gonzalez, R., and Navarro, A. (1991). Two beta-glucanase genes are clustered in *Bacillus polymyxa*: molecular cloning, expression, and sequence analysis of genes encoding a xylanase and an endo-beta-(1,3)-(1,4)-glucanase. *Journal of Bacteriology*. **173**, 7705-7710.

Goyal, A.K., and Eveleigh, D.E. (1996). Cloning, sequencing and analysis of the *ggh-A* gene encoding a 1,4-β-D-glucan glucohydrolase from *Microspora bispora*. *Gene*. **172**, 93-98.

Halstead, J.R., Vercoe, P., Gilbert, H.J., Davidson, K., and Hazlewood, G.P. (1999). A family 26 mannanase produced by Clostridium thermocellum as a component of the cellulosome contains a domain which is conserved in mannanases from anaerobic fungi. *Microbiology*. **145**, 3101-3108

Hamai, A., Morikawa, K., Horie, K., and Tokuyasu, K. (1989). Purification and characterization of hyaluronidase from *Streptococcus dysgalactiae*. *Agricultural and Biological Chemistry*. **53**, 2163-2168.

Hamai, A., Hashimoto, N., Mochizuki, H., Kato, F., Makiguchi, Y., Horie, K., and Suzuki, S. (1997). Two distinct chondroitin sulfate ABC lyases. An endoeliminase yielding tetrasaccharides and an exoeliminase preferentially acting on oligosaccharides. *The Journal of Biological Chemistry*. **272** (14), 9123-9130.

Hanahan, D. (1983). Studies on transformation of *Escherichia coli* with plasmids. *Journal of Molocular Biology*. **166** (4), 557-580.

Hansen, C.K. (1992). Fibronectin type III-like sequences and a new domain type in prokaryotic depolymerises with insoluble substrates. *FEBS Letters*. **305**, 91-96.

Hardingham, T.E., and Fosang, A.J. (1992). Proteoglycans: many forms and many functions. *The Federation of American Societies for Experimental Biology Journal*. **6** (3), 861-870.

Hashimoto, W., Kobayashi, E., Nankai, H., Sato, N., Miya, T., Kawai, S., and Murata, K. (1999). Unsaturated glucuronyl hydrolase of Bacillus sp. GL1: Novel enzyme prerequisite for metabolism of unsaturated oligosaccharides produced by polysaccharide lyases. *Archives of Biochemistry and Biophysics*. **368**, 367-374.

Hazlewood, G. P., Romaniec, M.P.M., Davidson, K., Grépinet, O., Béguin, P., Millet, J., Raynaud, O., and Aubert, J.-P. (1988). A catalogue of *Clostridium thermocellum* endoglucanase, β-glucosidase and xylanase genes cloned in *Escherichia coli*. *FEMS Microbiology Letters*. **51**, 231-236.

Hazlewood, G. P., and Glibert, H. J. (1993). Molecular biology of hemicellulases. In: *Hemicellulose and Hemicelulases*. M. P., Coughlan, & G. P., Hazlewood (Eds.). pp 103-126, Portland Press Ltd, London.

Hazlewood, G. P., and Glibert, H. J. (1998). Structure and function analysis of *Pseudomonas* plant cell wall hydrolases. *Biochemical Society Transactions*. **26**, 185-190.

Hazlewood, G. P., and Glibert, H. J. (1998). Structure and function analysis of *Pseudomonas* plant cell wall hydrolases. *Progress in Nucleic Acid Research and Molecular Biology*. **61**, 211-241.

Hegde, S.S., Kumar, A.R., Ganesh, K.N., Swaminathan, C.P., and Khan, M.I. (1998). Thermodynamics of ligand (substrate/end product) binding to endoxylanase from *Chainia* sp. (NCL-82-5-1): isothermal calorimetry and fluorescence titration studies. *Biochimica et Biophysica Acta*. **1388**, 93-100.

Heldin, P., Suzuki, M., Teder, P., and Pertoft, H. (1995). Chondroitin sulphate proteoglycan modulates the permeability of hyaluronan-containing coats around normal human mesothelial cells. *Journal of Cellular Physiology*. **165** (1), 54-61

Henrissat, B. (1991). A classification of glycosyl hydrolases based on amino acid sequence similarities. *Biochemical Journal*. **280**, 309-316.

Henrissat, B., Callebaut, I., Fabrega, S., Lehn, P., Mornon, J-P., and Davies, G.J. (1995). Conserved catalytic machinery and the prediction of a common fold for several families of glycosyl hydrolases. *Proceedings of the National Academy of Sciences, USA*. **92**, 7090-7094.

Henrissat, B., and Bairoch, A. (1996). Updating the sequence-based classification of glycosyl hydrolases. *Biochemical Journal*. **316**, 695-696.

Henrissat, B., and Davies, G.J. (1997). Structural and sequence-based classification of glycoside hydrolases. *Current Opinion in Structural Biology*. **7**, 637-644.

Henrissat, B. (1998). Glycosidase families. *Biochemical Society Transactions*. **26**, 153-156.

Henrissat, B., Teeri, T.T., and Warren, R.A.J. (1998). A scheme for designating enzymes that hydrolyse the polysaccharides in the cell walls of plants. *FEBS Letters*. **425**, 352-354.

Henrissat, B., and Coutinho, P. (2000). CAZy website: http://afmb.cnrs-mrs.fr/CAZY/.

Henrissat, B., and Davies, G.J. (2000). Glycoside hydrolases and glycosyltransferases. Families, modules, and implications for genomics. *Plant Physiology*. **124**, 1515-1519.

Henshaw, J., Bolam, D.N., Pires, V.M.R., Czjzek, M., Henrissat, B., Ferreira, L.M.A., Fontes, C.M.G.A. and Gilbert, H.J. (2004). The family 6 carbohydrate-binding module *Cm*CBM6-2 contains two ligand-binding sites with distinct specificities. *The Journal of Biological Chemistry*. In press.

Holmbeck, S., and Lerner, L. (1993). Separation of hyaluronan oligosaccharides by the use of anion-exchange HPLC. *Carbohydrate Research*. **239**, 239-244.

Homer, C., and Shulman, S.T. (1991). Clinical aspects of acute rheumatic fever. *Journal of rheumatology Supplement*. Review, **29**, 2-13.

Hong, T.-Y., Cheng, C-W., Huang, J.W., and Meng, M. (2002). Isolation and biochemical characterization of an endo-1,3-β-glucanase from *Streptomyces sioyaensis* containing a C-terminal family 6 carbohydrate-binding module that binds to 1,3-β-glucan. *Microbiology*. **148**, 1151-1159.

Hotez, P.J., Narasimhan, S., Haggerty, J., Milstone, L., Bhopale, V., Schad, G.A., and Richards, F.F. (1992). Hyaluronidase from infective *Ancylostoma* hookworm larvae and its possible function as a virulence factor in tissue invasion and in cutaneous larva migrans. *Infection and Immunity*. **60** (3), 1018-1023.

Hotez, P., Cappello, M., Hawdon, J., Beckers, C., and Sakanari, J. (1994). Hyaluronidases of the gastrointestinal invasive nematodes *Ancylostoma caninum* and *Anisakis simplex*: possible functions in the pathogenesis of human zoonoses. *Journal of Infectious Disease*. **170** (4), 918-926.

Hsiung, H.M., Mayne, N.G., and Becker, G.W. (1986). High-level expression, efficient secretion and folding of human growth-hormone in *Escherichia coli*. *Bio/Technology*. **4**, 1002-1013.

Huang, W., Boju, L., Tkalec, L., Su, H., Yang, H.O., Gunay, N.S., Linhardt, R.J., Kim, Y.S., Matte, A., and Cygler, M. (2001). Active site of chondroitin AC lyase revealed by the structure of enzyme-oligosaccharide complexes and mutagenesis. *Biochemistry*. **40** (8), 2359-2372.

Huang, W., Lunin, V.V., Li, Y., Suzuki, S., Sugiura, N., Miyazono, H., and Cygler, M. (2003). Crystal structure of *Proteus vulgaris* chondroitin sulfate ABC lyase I at 1.9A resolution. *Journal of Molecular Biology*. **328** (3), 623-634.

Huber, M., Trattnig, S., and Lintner, F. (2000). Anatomy, biochemistry and physiology of articular cartilage. *Investigative Radiology*. **35** (10), 573-580.

Humphry, D. R., Black, G. W. and Cummings, S. P. (2003). Reclassification of '*Pseudomonas fluorescens* subspecies *cellulosa*' NCIMB 10462 (Ueda *et al*, 1952) as *Cellvibrio japonicus* sp. nov. and revival of *Cellvibrio vulgaris* sp. nov.

nom. Rev. and *Cellvibrio fulvus* sp. nov. nom. rev. *International Journal of Systemic and Evolutionary Microbiology*. **53**: 393-400.

Hynes, W.L., and Ferretti, J.J. (1989). Sequence analysis and expression in *Escherichia coli* of the hyaluronidase gene of *Streptococcus pyogenes* bacteriophage H4489A. *Infection and Immunity*. **57** (2), 533-539.

Hynes, W.L., and Ferretti, J.J. (1994). Assays for hyaluronidase activity. *Methods in Enzymology*. **235**, 606-616.

Hynes, W.L., Hancock, L., and Ferretti, J.J. (1995). Analysis of a second bacteriophage hyaluronidase gene from *Streptococcus pyogenes*: evidence for a third hyaluronidase involved in extracellular enzymatic activity. *Infection and Immunity*. **63** (8), 3015-3020.

Hynes, W.L., and Walton, S.L. (2000). Hyaluronidases of Gram-positive bacteria. *FEMS Microbiology Letters*. **183**, 201-207.

Hynes, W.L., Dixon, A.R., Walton, S.L., and Aridgides, L.J. (2000). The extracellular hyaluronidase gene (*hylA*) of *Streptococcus pyogenes*. *FEMS Microbiology Letters*. **184**, 109-112.

Ikegami-Kawai, M., and Takahashi, T. (2002). Microanalysis of hyaluronan oligosaccharides by polyacrylamide gel electrophoresis and its application to assay of hyaluronidase activity. *Analytical Biochemistry*. **311**, 157-165.

Imanari, T., Toida, T., Koshiishi, I., and Toyoda, H. (1996). High-performance liquid chromatographic analysis of glycosaminoglycan-derived oligosaccharides. *Journal of Chromatography A*. **720**, 275-293.

Jenkins, J., Lo Leggio, L., Harris, G., and Pickersgill, R.W. (1995). β-glucosidase, β-galactosidase, family A cellulases, family F xylanases and two barley glycanases form a superfamily of enzymes with 8-fold β/α architecture and two conserved glutamates near the carboxy-terminal ends of β-strands four and seven. *FEBS Letters*. **362**, 281-285.

Jing, W., and DeAngelis, P.L. (2000). Dissection of the two transferase activities of the *Pasteurella multocida* hyaluronan synthase: two active sites exist in one polypeptide. *Glycobiology*. **10**, 883-889.

Johnson, D.R., and Kaplan, E.L. (1993). A review of the correlation of T-agglutination patterns and M-protein typing and opacity factor production in the identification of group A streptococci. *Journal of Medical Microbiology*. **38** (5), 311-315.

Johnson, P.E., Joshi, M.D., Tomme, P., Kilburn, D.G., and McIntosh, L.P. (1996). Structure of the N-terminal cellulose-binding domain of *Cellulomonas fimi* CenC determined by nuclear magnetic resonance spectroscopy. *Biochemistry*. **35**, 14381-14394.

Joseleau, J.P., Cartier, N., Chambat, G., Faik, A., and Ruel, K. (1992). Structural features and biological activity of xyloglucans from suspension-cultured plant cells. *Biochimie.* **74** (1), 81-88.

Joshi, M., Sidhu, G., Pot, I., Brayer, G.D., Withers, S.G., and McIntosh, L.P. (2000). Hydrogen bonding and catalysis: a novel explanation for how a single amino acid substitution can change the pH optimum of a glycosidase. *Journal of Molecular Biology.* **299**, 255-279.

Joy, M.B., Dodgson, K.S., Olavesen, A.H., and Gacesa, P. (1985). The purification and some properties of pig liver hyaluronidase. *Biochimica et Biophysica Acta.* **838**, 257-263.

Kanamaru, S., Leiman, P.G., Kostyuchenko, V.A., Chipman, P.R., Mesyanzhinov, V.V., Arisaka, F., and Rossmann, M.G. (2002). Structure of the cell-puncturing device of bacteriophage T4. *Nature.* **415**, 553-557.

Kapur, V., Topouzis, S., Majesky, M.W., Li, L.L., Hamrick, M.R., Hamill, R.J., Patti, J.M., and Musser, J.M. (1993). A conserved *streptococcus pyogenes* extracellular cysteine protease cleaves human fibronectin and degrades vitronectin. *Microbial Pathogenesis.* **15** (5), 327-346.

Karlstam, B., and Ljunglof, A. (1991). Purification and partial characterization of a novel hyaluronic acid-degrading enzyme from antarctic krill (*Euphausia superba*). *Polar Biology.* **11**, 501-507.

Kehoe, M.A., Kapur, V., Whatmore, A.M., and Musser, J.M. (1996). Horizontal gene transfer among group A streptococci: implications for pathogenesis and eipdemiology. *Trends in Microbiology.* **4** (11), 436-443.

Kellet, L.E., Poole, D.M., Ferreira, L.M.A., Durrant, A.J., Hazlewood, G.P., and Gilbert, H.J. (1990). Xylanase B and an arabinofuranosidase from *Pseudomonas fluorescens* subsp. *cellulosa* contain identical cellulose-binding domains and are encoded by adjacent genes. *Biochemical Journal.* **272**, 369-376.

Kemeny, D.M., Dalton, N., Lawrence, A.J., Pearce F.L., and Vernon, C.A. (1984). The purification and characterisation of hyaluronidase from the venom of the honey bee, *Apis mellifera. European Journal of Biochemistry.* **139**, 217-223.

Kjellén, L., and Lindahl, U. (1991). Proteoglycans: structures and interactions. *Annual Review of Biochemistry.* **60**, 443-475.

Kittur, F.S., Mangala, S.L., Rus'd, A.A., Kitaoka, M., Tsujibo, H., and Hayashi, K. (2003). Fusion of family 2b carbohydrate-binding module increases the catalytic activity of a xylanase from *Thermotoga maritima* to soluble xylan. *FEBS Letters.* **549**, 147-151.

Kizaki, K., Nakano, H., Nakano, T., Takahashi, T., Imai, K., and Hashizume, K. (2001). Expression of heparanase mRNA in bovine placenta during gestation. *Reproduction.* **121**, 573-580.

Knudson, W., Bartnik, E., and Knudson, C.B. (1993). Assembly of pericellular matrices by COS-7 cells transfected with CD44 lymphocyte-homing receptor genes. *Proceedings of the National Academy of Sciences, USA.* **90**, 4003-4007.

Knudson, C.B., and Knudson, W. (2001). Cartilage proteoglycans. *Cell and Development Biology.* **12**, 69-78.

Knudson, W., Chow, G., and Knudson, C.B. (2002). CD44-mediated uptake and degradation of hyaluronan. *Matrix Biology.* **21**, 15-23.

Kohorn, B. D. (2000). Plasma Membrane-Cell Wall Contacts. *Plant Physiology.* **124,** 31-38.

Kojima, T., Leone, C.W., Marchildon, G.A., Marcum, J.A., and Rosenberg, R.D. (1992). Isolation and characterization of heparan sulfate proteoglycans produced by cloned rat microvascular endothelial cells. *The Journal of Biological Chemistry.* **267** (7), 4859-4869.

Kojima, T., Shworak, N.W., and Rosenberg, R.D. (1992). Molecular cloning and expression of two distinct cDNA-encoding heparan-sulfate proteoglycan core proteins from a rat endothelial cell line. *The Journal of Biological Chemistry.* **267** (7), 4870-4877.

Kormos, J., Johnson, P.E, Brun, E., Tomme, P., McIntosh, L.P., Haynes, C.A., and Kilburn, D.G. (2000). Binding site analysis of cellulose binding domain CBD(N1) from endoglucanse C of *Cellulomonas fimi* by site-directed mutagenesis. *Biochemistry.* **39** (30), 8844-8852.

Koshiishi, I., Takenouchi, M., Hasegawa, T., and Imanari, T. (1998). Enzymatic method for the simultaneous determination of hyaluronan and chondroitin sulfates using high-performance liquid chromatography. *Analytical Biochemistry.* **265**, 49-54.

Kramer, R.Z., Vitagliano, L., Bella, J., Berisio, R., Mazzarella, L., Brodsky, B., Zagari, A., and Berman, H.M. (1998). X-ray crystallographic determination of a collagen-like peptide with the repeating sequence (Pro-Pro-Gly). *Journal of Molecular Biology.* **280**, 623-638.

Kramer, R.Z., Venugopal, M.G., Bella, J., Mayville, P., Brodsky, B., and Berman, H.M. (2000). Staggered molecular packing in crystals of a collagen-like peptide with a single charged pair. *Journal of Molecular Biology.* **301**, 1191-1205.

Kramer, R.Z., Bella, J., Brodsky, B., and Berman, H.M. (2001). The crystal and molecular structure of a collagen-like peptide with a biologically relevant sequence. *Journal of Molecular Biology.* **311**, 131-147.

Kreil, G. (1995). Hyaluronidases: a group of neglected enzymes. *Protein Science.* **4**, 1666-1669.

Kroon, P.A., Williamson, G., Fish, N.M., Archer, D.B., and Belshaw, N.J. (2000). A modular esterase from *Penicillium funiculosum* which releases ferulic acid from plant cell walls and binds crystalline cellulose contains a carbohydrate binding module. *European Journal Biochemistry*. **267**, 6740-6752.

Kuettner, K.E. (1992). Biochemistry of articular cartilage in health and disease. *Clinical Biochemistry*. **25** (3), 155-163.

Kuettner, K.E., and Thonar, E.J. (1998). Osteoarthritis: cartilage integrity and homeostasis. In: *Rheumatology*. J.H. Klippel and P.A. Dieppe (Eds.). Second Edition. pp 6.1-6.16.

Kuno, A., Kaneko, S., Ohtsuki, H., Ito, S., Fujimoto, Z., Mizuno, H., Hasegawa, T., Taira, K., Kusukabe, I., and Hayashi, K. (2000). Novel sugar-binding specificity of the type XIII xylan-binding domain of a family F/10 xylanase from *Streptomyces olivaceoviridis* E-86. *FEBS Letters*. **482**, 231-236.

Kuno, A., Shimizu, D., Kaneko, S., Koyama, Y., Yoshida, S., Kobayashi, H., Hayashi, K., Taira, K., and Kusukabe, I. (1998). PCR cloning and expression of the family F/10 xylanase gene from *Streptomyces olivaceoviridis* E-86. *Journal of Fermentation and Bioengineering*. **86**, 434-439.

Kussie, P.H., Hulmes, J.D., Ludwig, D.L., Patel, S., Navarro, E.C., Seddon, A.P., Giorgio, N.A., and Bohlen, P. (1999). Cloning and functional expression of a human heparanase gene. *Biochemical and Biophysical Research Communications*. **261** (1), 183-187.

Laemmli, U.K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature*. **227**, 680-685.

Lancefield, R.C. (1928). The antigenic complex of *Streptococcus haemolyticus*. *Journal of Experimental Medicine*. **47**, 91-103.

Lancefield, R.C. (1959). Persistence of type-specific antibodies in man following infection with group A streptococci. *Journal of Experimental Medicine*. **110** (2), 271-292.

Lancefield, R.C. (1962). Current knowledge of type-specific M antigens of group A streptococci. *Journal of Immunology*. **89**, 307-313.

Lancefield, R.C. (1969). Current problems in studies of Streptococci. *Journal of General Microbiology*. **55** (2), 161-163.

Lauder, R.M., Hucherby, T.N., and Nieduszynski, I.A. (2000). A fingerprinting method for chondroitin/dermatan sulfate and hyaluronan oligosaccharides. *Glycobiology*. **10** (4), 393-401.

Laurent, U.B., Fraser, J.R., Engstrom-Laurent, A., Reed, R.K., Dahl, L.B., and Laurent, T.C. (1992). Catabolism of hyaluronan in the knee joint of the rabbit. *Matrix*. **12** (2), 130-136.

Laurent, T.C., Laurent, U.B.G. and Fraser, J.R.E. (1996). The structure and function of hyaluronan: an overview. *Immunology and Cell Biology*. **74**, A1-A7.

Lehtiö, J., Sugiyama, J., Gustavsson, M., Fransson, L., Linder, M., and Teeri, T.T. (2003). The binding specificity and affinity determinants of family 1 and family 3 cellulose binding modules. *Proceedings of the National Academy of Sciences, USA*. **100** (2), 484-489.

Lemaire, M., Miras, I., Gounon, P., and Béguin, P. (1998). Identification of a region responsible for binding of the cell wall with the S-layer protein of *Clostridium thermocellum*. *Microbiology*. **144**, 211-217.

Leng, L., Zhu, A., Zhang, Z., Hurst, R., and Goldstein, J. (1998). Cloning, functional expression and purification of endo-beta-galactosidase from Flavobacterium keratolyticus. *Gene*. **222**, 187-194.

Lepperdinger, G., Strobl, B., and Kreil, G. (1998). *HYAL2*, a human gene expressed in many cells, encodes a lysosomal hyaluronidase with a novel type of specificity. *The Journal of Biological Chemistry*. **273** (35), 22466-22470.

Lerouge, P., O'Meill, M., Darvill, A.G., and Albersheim, P. (1993). Structural characterization of endo-glucanase-generated oligoglycosyl side chains of rhamnogalacturonan I. *Carbohydrate Research*. **243**, 359-371.

Leslie, A.W.G. (1992). Joint CCP4-ESF-EACMB. *Newsletter for Protein Crystallography*. **26**.

Lesley, J., Hyman, R., and Kincade, P.W. (1993). CD44 and its interaction with extracellular matrix. *Advances in Immunology*. **54**, 271-335.

Li, K., Azadi, P., Collins, R., Tolan, J., Kim, J.S., and Eriksson, K.-E.L. (2000). Relationships between activities of xylanases and xylan structures. *Enzyme and Microbial Technology*. **27**, 89-94.

Li, M-W., Yudin, A.I., Robertson, K.R., Cherr, G.N., and Overstreet, J.W. (2002). Importance of glycosylation and disulphide bonds in hyaluronidase activity of macaque sperm surface PH-20. *Journal of Andrology*. **23** (2), 211-219.

Lin, Y., Kimmel, L.H., Myles, D.G., and Primakoff, P. (1993). Molecular cloning of the human and monkey sperm surface protein PH-20. *Proceedings of the National Academy of Sciences, USA*. **90**, 10071-10075.

Lin, Y., Mahan, K., Lathrop, W.F., Myles, D.G., and Primakoff, P. (1994a). A hyaluronidase activity of the sperm plasma membrane protein PH-20 enables sperm to penetrate the cumulus cell layer surrounding the egg. *The Journal of Cell Biology*. **125** (5), 1157-1163.

Lin, B., Hollingshead, S.K., Coligan, J.E., Egan, M.L., Baker, J.R., and Pritchard, D.G. (1994b). Cloning and expression of the gene for group B streptococcal hyaluronate lyase. *The Journal of Biological Chemistry*. **269** (48), 30113-30116.

Lindahl, U., Kusche, M., Lidholt, K., and Oscarsson, L.G. (1989). Biosynthesis of heparin and heparan sulfate. *Annals of the New York Acadamey of Sciences*. **556**, 36-50.

Lindahl, U., Lidholt, K., Spillmann, D., and Kjellen, L. (1994). More to "heparin" than anticoagulation. *Thrombosis Research*. **75** (1), 1-32.

Linder, M., Mattinen, M.L., Kontteli, M., Lindeberg, G., Ståhlberg, J., Drakenberg, T., Reinikainen, T., Pettersson, G., and Annila, A. (1995). Identification of functionally important amino acids in the cellulose binding domain of *Trichoderma reesei* cellobiohydrolase I. *Protein Science*. **4**, 1056-1064.

Linder, M., Salovuori, I., Ruohonen, L., and Teeri, T.T. (1996). Characterization of a double cellulose-binding domain. Synergistic high affinity binding to crystalline cellulose. *The Journal of Biological Chemistry*. **271** (35):21268-21272.

Linder, M., and Teeri, T.T. (1997). The roles and function of cellulose-binding domains. *Journal of Biotechnology*. **57**, 15-28.

Linhardt, R.J., Galliher, P.M., and Cooney, C.L. (1986). Polysaccharide lyases. *Applied Biochemistry and Biotechnology*. **2** (2), 135-176.

Linn, S., Chan, T., Lipeski, L., Salyers, A.A. (1983). Isolation and characterization of two chondroitin lyases from *Bacteroides thetaiotaomicron*. *Journal of Bacteriology*. **156** (2), 859-866.

Linsenmayer, T.F. (1991). Collagen. In: *Cell Biology of Extracellular Matrix*. E.D. Hay (Ed.). Second edition. pp 7-40. Plenum Press, New York.

Little, E., Bork, P., and Doolittle, R.F. (1994). Tracing the spread of fibronectin type III domains in bacterial glycohydrolases. *Journal of Molecular Biology and Evolution*. **39**, 631-643.

Lohse, D.L., and Linhardt, R.J. (1992). Purification and characterization of heparin lyases from *Flavobacterium heparinum*. *The Journal of Biological Chemistry*. **267** (34), 24347-24355.

Lokeshwar, V.B., Lokeshwar, B.L., Pham, H.T., and Block, N.L. (1996). Association of elevated levels of hyaluronidase, a matrix-degrading enzyme, with prostate cancer progression. *Cancer Research*. **56** (3), 651-657.

Lu, G., Kochoumian, L., and King, T.P. (1995). Sequence identity and antigenic cross-reactivity of white face hornet venom allergen, also a hyaluronidase, with other proteins. *The Journal of Biological Chemistry*. **270** (9), 4457-4465.

Lunin, V.V., Li, Y., Linhardt, R.J., Miyazono, H., Kyogashima, M., Kaneko, T., Bell, A.W., and Cygler, M. (2004). High-resolution crystal structure of *Arthrobacter aurescens* chondroitin AC lyase: an enzyme-substrate complex defines the catalytic mechanism. *Journal of Molecular Biology*. **337** (2), 367-386.

Mahoney, D.J., Aplin, R.T., Calabro, A., Hascall, V.C., and Day, A.J. (2001). Novel methods for the preparation and characterization of hyaluronan oligosaccharides of defined length. *Glycobiology*. **11** (12), 1025-1033.

Marcum, J.A., Atha, D.H., Fritze, L.M., Nawroth, P., Stern, D., and Rosenberg, R.D. (1986). Cloned bovine aortic endothelial cells synthesize anticoagulantly active heparan sulphate proteoglycan. *The Journal of Biological Chemistry*. **261** (16), 7507-7517.

Marković-Housley, Z., Miglierini, G., Soldatova, L., Rizkallah, P.J., Muller, U., Schirmer, T. (2000). Crystal structure of hyaluronidase, a major allergen of bee venom. *Structure with Folding and Design*. **8** (10), 1025-1035.

Mattinen, M., Linder, M., Teleman, A., and Annila, A. (1997). Interaction between cellohexaose and cellulose-binding domains from *Trichoderma reesei* cellulases. *FEBS Letters*. **407**, 291-296.

McCann, M. C., Bush, M., Milioni, D., Sado, P., Stacey, N. J., Catchpole, G., Defernez, M., Carpita, N. C., Hofte, H., Ulvskov, P., Wilson, R. H., and Roberts, K. (2001). Approaches to understanding the functional architecture of the plant cell wall. *Phytochemistry*. **57**, 811-821.

McKie, V.A., Vincken, J.P., Voragen, A.G., van den Broek, L.A., Stimson, E., and Gilbert, H.J. (2001). A new family of rhamnogalacturonan lyases contains an enzyme that binds to cellulose. *Biochemical Journal*. **355** (Pt 1), 167-77.

McCarter, J.D., and Withers, S.G. (1994). Mechanisms of enzymatic glycoside hydrolysis. *Current Opinion in Structural Biology*. **4**, 885-892.

McClean, D. (1941). The capsulation of Streptococci and its relation to diffusion factor (Hyaluronidase) *The Journal of Pathology and Bacteriology*. **53**, 13-27.

McLean, B.W., Bray, M.R., Boraston, A.B., Gilkes, N.R., Haynes, C.A., and Kilburn, D.G. (2000). Analysis of binding of the family 2a carbohydrate-binding module from *Cellulomonas fimi* xylanase 10A to cellulose: specificity and identification of functionally important amino acid residues. *Protein Engineering*. **13** (11), 801-809.

McNeil, M., Darvill, A. G., Fry, S. C. & Albersheim, P. (1984) Structure and function of the primary cell walls of plants. *Annual Review of Biochemistry*. 53, 625-663.

Melrose, J., Tammi, M., and Smith, S. (2002). Visualisation of hyaluronan and hyaluronan-binding proteins within ovine vertebral cartilages using biotinylated

aggrecan G1-link complex and biotinylated hyaluronan oligosaccharides. *Histochemistry and Cell Biology*. **117**, 327-333.

Menasche, M., Dagonet, F., Ferrari, P., and Labat-Robert, J. (2001). Fibronectin in the vitreous body – distribution and possible functional role. *Pathologie Biologie*. **49**, 290-297.

Menzel, E.J., and Farr, C. (1998). Hyaluronidase and its substrate hyaluronan: biochemistry, biological activities and therapeutic uses. *Cancer Letters*. **131**, 3-11.

Meyer, K. Dubos, R., and Smyth, E.M. (1937). The hydrolysis of the polysaccharide acids of vitreous humor, of umbilical cord, and of streptococcus by the autolytic enzymes of pneumococcus. *The Journal of Biological Chemistry*. **118**, 71-78.

Meyer, K. (1971). Hyaluronidases. In: *The Enzymes*. Vol. 5. P.D. Boyer (Ed.). Third Edition, pp 307-319. Academic Press, London.

Miller, G.L. (1959). The use of dinitrosalicylic acid reagent for the determination of reducing sugar. *Analytical Chemistry*. **31**, 426-428.

Millward-Saddler, S.J., Poole, D.M., Henrissat, B., Hazlewood, G.P., Clarke, J.H., and Gilbert, H.J. (1994). Evidence for a general role for high-affinity non-catalytic cellulose-binding domains in microbial plant cell wall hydrolases. *Molecular Microbiology*. **11**, 375-382.

Millward-Saddler, S.J., Davidson, K., Hazlewood, G.P., Black, G.W., and Gilbert, H.J. (1995). Novel cellulose-binding domains, NodB homologues and conserved modular architecture in xylanases from the aerobic soil bacteria *Pseudomonas fluorescens* subsp. *cellulosa* and *Cellvibrio mixtus*. *Biochemical Journal*. **312**, 39-48.

Min, H., and Cowman, M.K. (1986). Combined alcian blue and silver staining of glycosaminoglycans in polyacrylamide gels: applications to electrophoretic analysis of molecular weight distribution. *Analytical Biochemistry*. **155**, 275-285.

Mio, K., and Stern, R. (2002). Inhibitors of the hyaluronidases. *Matrix Biology*. **21**, 31-37.

Mooney, S.D., Kollman, P.A., and Klein, T.E. (2001). Conformational preferences of substituted pralines in the collagen triple helix. *Biopolymers*. **64**, 63-71.

Moses, A.E., Wessels, M.R., Zalcman, K., Alberti, S., Natanson-Yaron, S., Menes, T., and Hanski, E. (1997). Relative contributions of hyaluronic acid capsule and M protein to virulence in a mucoid strain of the group A streptococcus. *Infection and Immunity*. **65**, 64-71.

Muckenschnabel, I., Bernhardt, G., Spruss, T., Dietl, B., and Buschauer, A. (1998). Quantitation of hyaluronidases by the Morgan-Elson reaction: comparison

of the enzyme activities in the plasma of tumor patients and healthy volunteers. *Cancer Letters*. **131**, 13-20.

Muir, H. (1995). The chondrocyte, architect of cartilage. Biomechanics, structure, function and molecular biology of cartilage matrix macromolecules. *Bioessays*. **12**, 1039-1048.

Nader, H.B., Kobayashi, E.Y., Chavante, S.F., Tersariol, I.L.S., Castro, R.A.B., Shinjo, S.K., Naggi, A., Torri, G., Casu, B., and Dietrich, C.P. (1999). New insights on the specificity of heparin and heparan sulphate lyases from *Flavobacterium heparinum* revealed by the use of synthetic derivatives of K5 polysaccharide from *E. coli* and 2-O-desulphated heparin. *Glycoconjugate Journal*. **16**, 265-270.

Nagy, T., Simpson, P.J., Williamson, M.P., Hazlewood, G.P., Gilbert, H.J., and Orosz, L. (1998). All three surface tryptophans in type IIa cellulose binding domains play a pivatol role in binding both soluble and insoluble ligands. *FEBS Letters*. **429**, 312-316.

Nakajima, M., Irimura, T., Di Ferrante, D., Di Ferrante, N., and Nicolson, G.L. (1983). Heparan sulfate degradation: relation to tumor invasive and metastatic properties of mouse B16 melanoma sublines. *Science*. **220** (4597), 611-613.

Nakatani, H. (2002). Monte Carlo simulation of hyaluronidase reaction involving hydrolysis, transglycosylation and condensation. *Biochemical Journal*. **365**, 701-705.

Nishimoto, M., Fushinobu, S., Miyanaga, A., Wakagi, T., Shoun, H., Sakka, K., Ohmiya, K., Nirasawa, S., Kitaoka, M., and Hayashi, K. (2004). Crystallization and preliminary X-ray analysis of xylanase B from *Clostridium stercorarium*. *Acta Crystallographica Section D Biological Crystallography*. **D60**, 342-343.

Nordberg, K.E., Bartonek-Roxå, E., and Holst, O. (1997). Cloning and sequence of a thermostable mutli-domain xylanase from the bacterium *Rhodothermus marinus*. *Biochimica et Biophysica Acta*. **1353** (2), 118-124.

Notenboom, V., Boraston, A.B., Kilburn, D.G., and Rose, D.R. (2001). Crystal structures of the family 9 carbohydrate-binding module from *Thermotoga maritima* xylanase 10A in native and ligand-bound forms. *Biochemistry*. **40** (21), 6248-6256.

Notenboom, V., Boraston, A.B., Williams, S.J., Kilburn, D.G., and Rose, D.R. (2002). High-resolution crystal structures of the lectin-like xylan binding domain from *Streptomyces lividans* xylanase 10A with bound substrates reveal a novel mode of xylan binding. *Biochemistry*. **41**, 4246-4254.

Oettl, M., Hoechstetter, J., Asen, I., Bernhardt, G., and Buschauer, A. (2003). Comparative characterization of bovine testicular hyaluronidase and a hyaluronate lyase from *Streptococcus agalactiae* in pharmaceutical preparations. *European Journal of Pharmaceutical Sciences*. **18**, 267-277.

Okazaki, F., Tamaru, Y., Hashikawa, S., Li, Y.T., and Araki, T. (2002). Novel carbohydrate-binding module of beta-1,3-xylanase from a marine bacterium, *Alcaligenes* sp. strain XY-234. *Journal of Bacteriology*. **184** (9), 2399-2403.

Pace, C.N., Vadjos, F., Fee, L., Grimsley, G., and Gray, T. (1995). How to measure and predict the molar absorption coefficient of a protein. *Protein Science*. **4**, 2411-2423.

Pauly, M., Albersheim, P., Darvill, A., and York, W. S. (1999). Molecular domains of the cellulose/xyloglucan network in the cell walls of higher plants. *The Plant Journal*. **20** (6), 629-639.

Payan, E., Jouzeau, J.Y., Lapicque, F., Muller, N., and Netter, P. (1993). Hyaluronidase degradation of hyaluronic acid from different sources: influence of the hydrolysis conditions on the production and the relative proportions of tetra- and hexasaccharide produced. *The International Journal of Biochemistry*. **25** (3), 325-329.

Pell, G., Williamson, M.P., Walters, C., Du, H., Gilbert, H.J., and Bolam. D.N. (2003). Importance of hydrophobic and polar residues in ligand binding in the family 15 carbohydrate-binding module from *Cellvibrio japonicus* Xyn10C. *Biochemistry*. **42** (31), 9316-9323.

Pilot, I. (1944). The role of mucoid haemolytic streptococci in rheumatic fever and arthritis. *Proceedings of the Central Society for Clinical Research*. **17**, 70-71.

Pires, V.M.R., Henshaw, J., Prates, J.A.M., Bolam, D., Ferreira, L.M.A., Fontes, C.M.G.A., Henrissat, B., Planas, A., Gilbert, H.J., and Czjzek, M. (2004). The crystal structure of the family 6 carbohydrate-binding module from *Cellvibrio mixtus* lichenase 5A in complex with oligosaccharides reveals two distinct binding sites with different ligand specificities. *The Journal of Biological Chemistry*. In press.

Polansky, J.R., Toole, B.P., and Gross, J. (1974). Brain hyaluronidase: changes in activity during chick development. *Science*. **183** (127), 862-864.

Polissi, A., Pontiggia, A., Feger, G., Altieri, M., Mottl, H., Ferrari, L., and Simon, D. (1998). Large-scale identification of virulence genes from *Streptococcus pneumoniae*. *Infection and Immunity*. **66** (12), 5620-5629.

Ponyi, T., Szabó, L., Nagy, T., Orosz, L., Simpson, P.J., Williamson, M.P., and Gilbert, H.J. (2000). Trp22, Trp24, and Tyr8 play a pivotal role in the binding of the family 10 cellulose-binding module from *Pseudomonas* xylanase A to insoluble ligands. *Biochemistry*. **39** (5), 985-991.

Prescott, J.F., Fernandez, A.S., Nicholson, V.M., Patterson, M.C., Yager, J.A., Viel, L., and Perkins, G. 1996). Use of a virulence-associated protein based enzyme-linked immunosorbent assay for *Rhodococcus equi* serology in horses. *Equine Veterinary Journal*. **28** (5), 344-349.

Price, K.N., Tuinman, A., Baker, D.C., Chisena, C., and Cysyk, R.L. (1997). Isolation and characterization by electrospray-ionization mass spectrometry and high-performance anion-exchange chromatography of oligosaccharides derived from hyaluronic acid by hyaluronate lyase digestion: observation of some heretofore unobserved oligosaccharides that contain an odd number of units. *Carbohydrate Research*. **303**, 303-311.

Primakoff, P., Cowan, A., Hyatt, H., Tredick-Kline, J., and Myles, D.G. (1988). Purification of the guinea pig sperm PH-20 antigen and detection of a site-specific endoproteolytic activity in sperm preparations that cleaves PH-20 into two disulphide-linked fragments. *Biology of Reproduction*. **38**, 921-934.

Pritchard, D.G., Lin, B., Willingham, T.R., and Baker, J.R. (1994). Characterization of the group B streptococcal hyaluronate lyase. *Archives of Biochemistry and Biophysics*. **315** (2), 431-437.

Prockop, D.J, and Kivirikko, K.I. (1995). Collagens: molecular biology, diseases, and potentials for therapy. *Annual Review of Biochemistry*. **64**, 403-434.

Proft, T., Moffatt, S.L., Berkahn, C.J., and Fraser, J.D. (1999). Identification and characterization of novel superantigens from *Streptococcus pyogenes*. *Journal of Experimental Medicine*. **189** (1), 89-101.

Raghothama, S., Simpson, P.J., Szabó, L., Nagy, T., Gilbert, H.J., and Williamson, M.P. (2000). Solution structure of the CBM10 cellulose binding module from Pseudomonas xylanase A. *Biochemistry*. **39** (5), 978-984.

Rasmussen, M., Edén, A., and Björck, L. (2000). SclA, a novel collagen-like surface protein of *Streptococcus pyogenes*. *Infection and Immunity*. **68** (11), 6370-6377.

Reitinger, S., Müllegger, J., and Lepperdinger, G. (2001). *Xenopus* kidney hyaluronidase-1 (XKH1), a novel type of membrane-bound hyaluronidase solely degrades hyaluronan at neutral pH. *FEBS Letters*. **505**, 213-216.

Roberson, E., and Firestone, M. (1992). Relationship between desiccation and exopolysaccharide production in soil *Pseudomonas* sp. *Applied Environmental Microbiology*. **58**, 1284-91.

Roberts, I.S., Saunders, F.K., and Boulnois, G.J. (1989). Bacterial capsules and interactions with complement and phagocytes. *Biochemical Society Transactions*. **17**, 462-464.

Roberts, I.S. (1996). The biochemistry and genetics of capsular polysaccharide production in bacteria. *Annual Review of Microbiology*. **50**, 285-315.

Rotta, J., Krause, R.M., Lancefield, R.C., Everly, W., and Lackland, H. (1971). New approaches for the laboratory recognition of M types of group A streptococci. *The Journal of Experimental Medicine.* **134** (5), 1298-1315.

Roughley, P.J., and Lee, E.R. (1994). Cartilage proteoglycans: structure and function. *Microscopy Research and Technique.* **28**, 385-397.

Rye, C. S. and Withers, S. G. (2000). Glycosidase mechanisms. *Current Opinion in Chemical Biology.* **4**, 573-580.

Saitoh, H., Takagaki, K., Majima, M., Nakamura, T., Matsuki, A., Kasai, M., Narita, H., and Endo, M. (1995). Enzymatic reconstruction of glycosaminoglycan oligosaccharide chains using the transglycosylation reaction of bovine testicular hyaluronidase. *The Journal of Biological Chemistry.* **270** (8), 3741-3747.

Sakka, K., Nakanishi, M., Sogabe, M., Arai, T., Ohara, H., Tanaka, A., Kimura, T., Ohmiya, K. (2003). Isothermal titration calorimetric studies on the binding of a family 6 carbohydrate-binding module of *Clostridium thermocellum* xynA with xlylooligosaccharides. *Bioscience, Biotechnology and Biochemistry.* **67** (2), 406-409.

Sambrook, J., and Russell, D.W. (2001). Molecular Cloning: A Laboratory Maunal, 3rd Edition, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Schärpf, M., Connelly, G.P., Lee, G.M., Boraston, A.B., Warren, R.A.J., and McIntosh, L.P. (2002). Site-specific characterization of the association of xylooligosaccharides with the CBM13 lectin-like xylan binding domain from *Streptomyces lividans* xylanase 10A by NMR spectroscopy. *Biochemistry.* **41**, 4255-4263.

Schrager, H.M., Alberti, S., Cywes, C., Dougherty, G.J., and Wessels, M.R. (1998). Hyaluronic acid capsule modulates M protein-mediated adherence and acts as a ligand for attachment of group A streptococcus to CD44 on human kerayinocytes. *Journal of Clinical Investigation.* **101** (8), 1708-1716.

Schwartz, D.M., Shuster, S., Jumper, M.D., Chang, A., and Stern, R. (1996). Human vitreous hyaluronidase: isolation and characterization. *Current Eye Research.* **15** (12), 1156-1162.

Scott, J.E. (1995). Extracellular matrix, supramolecular organisation and shape. *Journal of Anatomy.* **187**, 259-269.

Scott, J.E., and Heatly, F. (1996). Secondary structures in water of chondroitin-4-sulphate and dermatan sulphate. Implications in the formation of tertiary structures. *Biochemical Society Transactions.* **24**, 98S.

Scott, J.E., and Heatly, F. (1999). Hyaluronan forms specific stable tertiary structures in aqueous solution: a ^{13}C NMR study. *Proceedings of the National Academy of Sciences, USA.* **96**, 4850-4855.

Scott, J.E., Thomlinson, A.M., and Prehm, P. (2003). Supramolecular organization in streptococcal pericellular capsules is based on hyaluronan teriary structures. *Experimental Cell Research*. **285**, 1-8.

Sengupta, S., Jana, M.L., Sengupta, D., and Naskar, A.K. (2000). A note on the estimation of microbial glycosidase activities by dinitrosalicylic acid reagent. *Applied Microbiology and Biotechnology*. **53**, 732-735.

Shimizu, M.T., Jorge, A.O., Unterkircher, C.S., Fantinato, V., and Paula, C.R. (1995). Hyaluronidase and chondroitin sulphatase production by different species of Candida. *Journal of Medical and Veterinary Mycology*. **33** (1), 27-31.

Shimizu, M.T., Almeida, N.Q., Fantinato, V., and Unterkircher, C.S. (1996). Studies on hyaluronidase, chondroitin sulphatase, proteinase and phospholipase secreted by Candida species. *Mycoses*. **39** (5-6), 161-167.

Shimizu, T., Ontani, K., Hirakawa, H., Ohshima, K., Yamashita, A., Shiba, T., Ogasawara, N., Haltori, M., Kuhara, S., and Hayashi, H. (2002). Complete genome sequence of *Clostridium perfringens*, an anaerobic flesh-eater. *Proceedings of the National Academy of Sciences*. **99** (2), 996-1001.

Shoham, Y., Lamed, R., and Bayer, E.A. (1999). The cellulosome concept as an efficient microbial strategy for the degradation of insoluble polysaccharides. *Trends in Microbiology*. 7 (7), 275-281.

Shriver, Z., Hu, Y., Pojasek, K., and Sasisekharan, R. (1998). Heparinase II from *Flavobacterium heparinum*. Role of cysteine in enzymatic activity as probed by chemical modification and site-directed mutagenesis. *The Journal of Biological Chemistry*. **273** (36), 22904-22912.

Shuttleworth, T.L., Wilson, M.D., Wicklow, B.A., Wilkins, J.A., and Triggs-Raine, B.L. (2002). Characterization of the murine hyaluronidase gene region reveals complex organization and cotranscription of *Hyal1* with downstream genes, *Fus2* and *Hyal3*. *The Journal of Biological Chemistry*. **277** (25), 23008-23018.

Simpson, H.D., and Barras, F. (1999). Functional analysis of the carbohydrate-binding domains of *Erwinia chrysanthemi* Cel5 (endoglucanase Z) and an *Escherichia coli* putative chitinase. *Journal of Bacteriology*. **181**, 4611-4616.

Simpson, P.J., Bolam, D.N., Cooper, A., Ciruela, A., Hazlewood, G.P., Gilbert, H.J., and Williamson, M.P. (1999). A family Iib xylan binding domain has a similar secondary structure to a homologous family Iia cellulose binding domain but different ligand specificity. *Structure*. **7**, 853-864.

Simpson, P.J., Xie, H., Bolam, D.N., Gilbert, H.J., and Williamson, M.P. (2000). The structural basis for the ligand specificity of family 2 carbohydrate-binding modules. *The Journal of Biological Chemistry*. **275** (52), 41137-41142.

Simpson, P.J., Jamieson, S.J., Abou-Hachem, M., Nordberg Karlsson, E., Gilbert, H.J., Holst, O., and Williamson, M.P. (2002). The solution structure of the CBM4-2 carbohydrate binding module from a thermostable *Rhodothermus marinus* xylanase. *Biochemistry*. **41**, 5712-5719.

Sinnott, M.L. (1990). Catalytic mechanisms of enzymic glycosyl transfer. *Chemistry Review*. **90**, 1171-1202.

Sorimachi, K., Le Gal-Coeffet, M.F., Williamson, G., Archer, D.B., and Williamson, M.P. (1997). Solution structure of the granular starch binding domain of *Aspergillus niger* glucoamylase bound to beta-cyclodextrin. *Structure*. **5** (5), 647-661.

Spezio, M., Wilson, D.B., and Karplus, P.A. (1993). Crystal structure of the catalytic domain of a thermophilic endocellulase. *Biochemistry*. **32**, 9906-9916.

Srisodsuk, M., Reinikainen, T., Penttilä, M., and Teeri, T.T. (1993). Role of the interdomain linker peptide of *Trichoderma reesei* cellobiohydrolase I in its interaction with crystalline cellulose. *The Journal of Biological Chemistry*. **268**, 20756-20761.

Stern, M., and Stern, R. (1992). A collagenous sequence in a prokaryotic hyaluronidase. *Molecular Biology and Evolution*. **9** (6), 1179-1180.

Stevens, D.L. (1992). Invasive group A streptococcus infections. *Clinical Infectious Disease*. **14** (1), 2-11. Review.

Stevens, D.L. Invasive group A streptococcal infections: the past, present and future. (1994). *Pediatrics and Infectious Disease Journal*. **13** (6), 561-5666. Review.

Stoll, D., Boraston, A., Stalbrand, H., McLean, B.W., Kilburn, D.G., and Warren, R.A. (2000). Mannanase Man26A from *Cellulomonas fimi* has a mannan-binding module. *FEMS Microbiology Letters*. **83** (2), 265-269.

Stollerman, G.H. (1996). The nature of rheumatogenic streptococci. *Mt. Sinai Journal of Medicine*. **63**, 144-158.

Stoolmiller, A.C., and Dorfman, A. (1969). The biosynthesis of hyaluronic acid by Streptococcus. *The Journal of Biological Chemistry*. **244** (2), 236-246.

Studier, F.W., and Moffat, B.A. (1986). Use of bacteriophage T7 RNA polymerase to direct the selective high-level expression of cloned genes. *Journal of Molecular Biology*. **189**, 113-130.

Su, H., Blain, F., Musil, R.A., Zimmermann, J.J.F., Gu, K., and Bennett, D.C. (1996). Isolation and expression in *Escherichia coli* of *hepB* and *hepC*, genes coding for the glycosaminoglycan-degrading enzymes Heparinase II and Heparinase III, respectively, from *Flavobacterium heparinum*. *Applied and Environmental Microbiology*. **62** (8), 2723-2734.

Sugahara, K., Yamada, S., Sugiura, M., Takeda, K., Yuen, R., Khoo, H.E., and Poh, C.H. (1992). Identification of the reaction products of the purified hyaluronidase from stonefish (*Synanceja horrida*) venom. *Biochemical Journal.* **283**, 99-104.

Sun, J.L., Sakka, K., Karita, S., Kimura, T., and Ohmiya, K. (1998). Adsorption of *Clostridium stercorarium* xylanase A to insoluble xylan and the importance of the CBDs to xylan hydrolysis. *Journal of Fermentation and Bioengineering.* **85**, 63-68.

Sunna, A., and Antranikian, G. (1997). Xylanolytic enzymes from fungi and bacteria. *Critical Reviews in Biotechnology.* **17** (1), 39-67.

Sunna, A., Gibbs, M.D., and Bergquist, P.L. (2000). The thermostabilizing domain, XynA, of *Caldibacillus cellulovorans* xylanase is a xylan binding domain. *Biochemical Journal.* **346**, 583-586.

Sunna, A., Gibbs, M.D., and Bergquist, P.L. (2001). Identification of novel β-mannan- and β-glucan-binding modules: evidence for a superfamily of carbohydrate-binding modules. *Biochemical Journal.* **356**, 791-798.

Sutherland, I.W. (1995). Polysaccharide lyases. *FEMS Microbiology Reviews.* **16**, 323-347.

Suzuki, A., Toyoda, H., Toida, T., and Imanari, T. (2001). Preparation and inhibitory activity on hyaluronidase of fully *O*-sulfated hyalruo-oligosaccharides. *Glycobiology.* **11** (1), 57-64.

Szabó, L., Jamal, S., Xie, H., Charnock, S.J., Bolam, D.N., Gilbert, H.J., and Davies, G.J. (2001). Structure of a family 15 carbohydrate-binding module in complex with xylopentaose. Evidence that xylan binds in an approximate 3-fold helical conformation. *The Journal of Biological Chemistry.* **276** (52), 49061-49065.

Takagaki, K., Nakamura, T., Izumi, J., Saitoh, H., Endo, M., Kojima, K., Kato, I., and Majima, M. (1994). Characterization of hydrolysis and transglycosylation by testicular hyaluronidase using ion-spray mass spectrometry. *Biochemistry.* **33**, (21), 6503-6507.

Takagaki, K., Ishido, K., Kakizaki, I., Iwafune, M., and Endo, M. (2002). Carriers for enzymatic attachment of glycosaminoglycan chains to peptide. *Biochemical and Biophysical Research Communications.* **293**, 220-224.

Takazono, I., and Tanaka, Y. (1984). Quantitative analysis of hyaluronic acid by high-performance liquid chromatography of streptomyces hyaluronidase digests. *Journal of Chromatography.* **288** (1), 167-176.

Takeo, K. (1984). Affinity electrophoresis: principles and applications. *Electrophoresis.* **5**, 187-195.

Tan, Y.H., and Bowness, J.M. (1968a). Canine submandibular-gland hyaluronidase. Identification and subcellular distribution. *Biochemical Journal.* **110** (1), 9-17.

Tan, Y.H., and Bowness, J.M. (1968b). Canine submandibular-gland hyaluronidase. Purification and properties. *Biochemical Journal.* **110** (1), 19-25.

Tawada, A., Masa, T., Oonuki, Y., Watanabe, A., Matsuzaki, Y., and Asari, A. (2002). Large-scale preparation, purification, and characterization of hyaluronan oligosaccharides from 4-mers to 52-mers. *Glycobiology.* **12** (7), 421-426.

Teeri, T.T. (1997). Crystaline cellulose degradation: new insight into the function of cellobiohydrolases. *TIBTECH.* **15**, 160-167.

Thibodeau, G.A., and Patton, K.T. (1992). The human body in health and disease. D. Allen, L.J. Edwards (Eds.). Mosby Year Book Inc.

Thomas, J. R., McNeil, M., Darvill, A. G., and Albersheim, P. (1987). Structure of Plant Cell Walls. XIX. Isolation and characterisation of wall polysaccharides from suspension-cultured Douglas fir cells. *Plant Physiology.* **83,** 659-671.

Timoney, J.F., Pesante, L., and Ernst, C. (1982). Hyaluronidase associated with a temperate bacteriophage of *Streptococcus equi. Microbiology.* 145-146.

Tkalec, A.L., Fink, D., Blain, F., Zhang-Sun, G., Laliberte, M., Bennett, D.C., Gu, K., Zimmermann, J.J., and Su, H. (2000). Isolation and expression in *Escherichia coli* of cslA and cslB, genes coding for the chondroitin sulfate-degrading enzymes chondroitinase AC and chondroitinase B, respectively, from *Flavobacterium heparinum. Applied Environmental Microbiology.* **66** (1), 29-35.

Tomme, P., van Tilbeurgh, H., Petterson, G., van Damme, J., Vandekerckhove, J., Knowles, J., Teeri, T., and Claeyssens, M. (1988). Studies of the cellulolytic system of *Trichoderma reesei* QM 9414. Analysis of domain function in two cellobiohydrolases by limited proteolysis. *European Journal of Biochemistry.* **170**, 575-581.

Tomme, P., Warren, R.A.J., Miller, R.C., Kilburn, D.G., and Gilkes, N.R. (1995a). Cellulose-binding domains – classification and properties. *ACS Symposium Series.* **618**, 142-162.

Tomme, P., Driver, D.P., Amandoron, E.A., Miller, JR., R.C., Warren, R.A.J., and Kilburn, D.G. (1995b). Comparison of a fungal (family I) and bacterial (family II) cellulose-binding domain. *Journal of Bacteriology.* **177** (15), 4356-4363.

Tomme, P., Warren, R.A.J., and Gilkes, N.R. (1995c). Cellulose hydrolysis by bacteria and fungi. *Advances in Microbial Physiology.* **37**, 1-81.

Tomme, P., Creagh, A.L., Kilburn, D.G. and Haynes, C.A. (1996). Interaction of polysaccharides with the N-terminal cellulose-binding domain of *Cellulomonas*

fimi CenC. 1. binding specificity and calorimetric analysis. *Biochemistry*. **35**, 13885-13894.

Tomme, P., Boraston, A., McLean, B., Kormos, J., Creagh, A.L., Sturch, K., Gilkes, N.R., Haynes, C.A., Warren, R.A.J., and Kilburn, D.G. (1998). Characterization and affinity applications of cellulose-binding domains. *Journal of Chromatography B*. **715**, 283-296.

Tomme, P., Boraston, A., Kormos, J.M., Warren, R.A.J. and Kilburn, D.G. (2000). Affinity electrophoresis for the identification and characterisation of soluble sugar binding by carbohydrate-binding modules. *Enzyme and Microbial Technology*. **27**, 453-458.

Tormo, J., Lamed, R., Chirino, A.J., Morag, E., Bayer, E.A., Shoham, Y., and Steitz, T.A. (1996). Crystal structure of a bacterial family-III cellulose-binding domain: a general mechanism for attachment to cellulose. *EMBO Journal*. **15** (21), 5739-5751.

Triggs-Raine, B., Salo, T.J., Zhang, H., Wicklow, B.A., and Natowicz, M.R. (1999). Mutations in *HYAL1*, a member of a tandemly distributed multigene family encoding disparate hyaluronidase activities, cause a newly described lysosomal disorder, mucopolysaccharidosis IX. *Proceedings of the National Academy of Sciences, USA*. **96**, 6296-6300.

Tull, D., and Withers, S.G. (1994). Mechanisms of cellulases and xylanases: a detailed kinetic study of the exo-β-1,4-glycanase from *Cellulomonas fimi*. *Biochemistry*. **33**, 6363-6370.

Usha, R., and Ramasami, T. (2004). The effects of urea and *n*-propanol on collagen denaturation: using DSC, circular dicroism and viscosity. *Thermochimica Acta*. **409** (2), 201-206.

van der Rest, M., and Garrone, R. (1991). Collagen family of proteins. *The Federation of American Societies for Experimental Biology Journal*. **5** (13), 2814-2823.

van Tilbeurgh, H., Tomme, P., Claeyssens, M., Bhikhabhai, R., and Petersson, G. (1986). Limited proteolysis of the cellobiohydrolase I from *Trichoderma reesei*: separation of functional domains. *FEBS Letters*. **204**, 223-227.

Varner, J. E. and Lin, L-S. (1989). Plant Cell Wall Architecture. *Cell*. **56**, 231-239.

Vercruysse, K.P., Ziebell, M.R., and Prestwich, G.D. (1999). Control of enzymatic degradation of hyaluronan by divalent cations. *Carbohydrate Research*. **318**, 26-37.

Vlodavsky, I., Friedmann, Y., Elkin, M., Aingorn, H., Atzmon, R., Ishai-Michaeli, R., Bitan, M., Pappo, O., Peretz, O., Michal, I., Spector, L., and Pecker, I. (1999).

Mammalian heparanase: gene cloning, expression and function in tumor progression and metastasis. *Nature Medicine.* **5** (7), 793-802.

Vocadlo, D.J., and Withers, S.G. (2000). Identification of active site residues in glycosidases by use of tandem mass spectrometry. *Methods in Molecular Biology.* **146**, 203-222.

Volpi, N., and Maccari, F. (2003). Purification and characterization of hyaluronic acid from the mollusc bivalve *Mytilus galloprovincialis. Biochimie.* **85**, 619-625.

Vrielink, A., Rüger, W., Driessen, H.P.C., and Freemont, P.S. (1994). Crystal structure of the DNA modifying enzyme β-glucosyltransferase in the presence and absence of the substrate uridine diphosphoglucose. *The EMBO Journal.* **13** (15), 3413-3422.

Wagner, P.L. and Waldor, M.K. (2002). Bacteriophage control of bacterial virulence. *Infection and Immunity.* **70** (8), 3985-3993.

Wang, Y., Slade, M.B., Gooley, A.A., Atwell, B.J., and Williams, K.L. (2001). Cellulose-binding modules from extracellular matrix proteins of *Dictyostelium discoideum* stalk and sheath. *European Journal of Biochemistry.* **268** (15), 4334-4345.

Warren, R.A.J. (1996). Microbial hydrolysis of polysaccharides. *Annual Review of Microbiology.* **50,** 183-212.

Weis, W.I., and Drickamer, K. (1996). Structural basis of lectin-carbohydrate recognition. *Annual Review of Biochemistry.* **65**, 441-473.

Wessels, M.R., Moses, A.E., Goldberg, J.B., and DiCesare, T.J. (1991). Hyaluronic acid capsule is a virulence factor for mucoid group A streptococci. *Proceedings of the National Academy of Sciences, USA.* **88**, 8317-8321.

Wessels, M.R., and Bronze, M.S. (1994). Critical role of the group A streptococcal capsule in pharyngeal colonization and infection in mice. *Proceedings of the National Academy of Science U S A.* **91** (25), 12238-42.

Wessels, M.R., Goldberg, J.B., Moses, A.E., and DiCesare, T.J. (1994). Effects on virulence of mutations in a locus essential for hyaluronic acid capsule expression in group A streptococci. *Infection and Immunity.* **62** (2), 433-441.

Westfall, B., Sitaraman, K., Lee, J.E., Borman, J., and Rashtchian, A. (1999). Platinum® Pfx DNA polymerase for high-fidelity PCR. *Focus.* **21** (2), 46-48.

Whatmore, A.M., and Kehoe, M.A. (1994). Horizontal gene transfer in the evolution of group A streptococcal emm-like genes: gene mosaics and variation in Vir regulons. *Molecular Microbiology.* **11** (2), 363-374.

Whatmore, A.M., Kapur, V., Sullivan, D.J., Musser, J.M and Kehoe, M.A. (1994). Non-congruent relationships between variation in emm gene sequences and the

population genetic structure of group A streptococci. *Molecular Microbiology*. **14** (4), 619-631.

Whitnack, E., Bisno, A.L., and Beachey, E.H. (1981). Hyaluronate capsule prevents attachment of group A streptococci to mouse peritoneal macrophages. *Infection and Immunity*. **31**, 985-991.

Williamson, G., Kroon, P.A., and Faulds, C.B. (1998). Hairy plant polysaccharides: a close shave with microbial esterases. *Microbiology*. **144**, 2011-2023.

Wilson, J.J., Matsushita, O., Okabe, A., and Sakon, J. (2003). A bacterial collagen-binding domain with novel calcium-binding motif controls domain orientation. *The EMBO Journal*. **22** (8), 1743-1752.

Winterhalter, C., Heinrich, P., Candussio, A., Wich, G., and Liebl, W. (1995). Identification of a novel cellulose binding domain within the multidomain 120 kDa xylanase XynA of the hyperthermophilic bacterium *Thermotoga maritime*. *Molecular Microbiology*. **15**, 431-444.

Wojtaszek, P. (2000). Genes and plant cell walls: a difficult relationship. *Biology Review*. **75**, 437-475.

Wyatt, S.E., and Carpita, N.C. (1993). The plant cytoskeleton-cell wall continuum. *Trends in Cell Biology*. **3**, 413-417.

Xie, H., Bolam, D.N., Nagy, T., Szabó, L., Cooper, A., Simpson, P.J., Lakey, J.H., Williamson, M.P., and Gilbert, H.J. (2001a). Role of hydrogen bonding in the interaction between a xylan binding module and xylan. *Biochemistry*. **40**, 5700-5707.

Xie, H., Gilbert, H.J., Charnock, S.J., Davies, G.J., Williamson, M.P., Simpson, P.J., Raghothama, S., Fontes, C.M.G.A., Dias, F.M.V., Ferreira, L.M.A., and Bolam, D.N. (2001b). *Clostridium thermocellum* Xyn10B carbohydrate-binding module 22-2: the role of conserved amino acids in ligand binding. *Biochemistry*. **40**, 9167-9176.

Xu, G.Y., Ong, E., Gilkes, N.R., Kilburn, D.G., Muhindaram, D.R., Harris-Brandts, M., Carver, J.P., Kay, L.E., and Harvey, T.S. (1995). Solution structure of a cellulose-binding domain from *Cellulomonas fimi* by nuclear magnetic resonance spectroscopy. *Biochemistry*. **34**, 6993-7009.

Xu, Y., Keene, D.R., Bujnicki, J.M., Höök, M., and Lukomski, S. (2002). Streptococcal Scl1 and Scl2 proteins form collogen-like triple helices. *The Journal of Biological Chemistry*. **277** (30), 27312-27318.

Yague, E., Beguin, P., Aubert, J.P. (1990). Nucleotide sequence and deletion analysis of the cellulase-encoding gene celH of *Clostridium thermocellum*. *Gene*. **89** (1), 61-67.

Yamada, M., Hasegawa, E., and Kanamori, M. (1977). Purification of hyaluronidase from human placenta. *Journal of Biochemistry* (Tokyo). **81** (2), 485-494.

Yamagishi, K., Suzuki, K., Imai, K., Mochizuki, H., Morikawa, K., Kyogashima, M., Kimata, K., and Watanabe, H. (2003). Purification, characterization, and molecular cloning of a novel keratan sulphate hydrolase, endo-β-*N*-acetylglucosaminidase, from *Bacillus circulans*. *The Journal of Biological Chemistry*. **278** (28), 25766-25772.

Yoshida, M., Itano, N. Yamada, Y., and Kimata, K. (2000). *In vitro* synthesis of hyaluronan by a single protein derived from mouse HAS1 gene and characterization of amino acid residues essential for the activity. *The Journal of Biological Chemistry*. **275** (1), 497-506.

Yuki, H., and Fishman, W.H. (1963). Purification and characterization of leech hyaluronic acid-endo-beta-glucuronidase. *The Journal of Biological Chemistry*. **238**, 1877-1879.

Zechel, D.L., and Withers, S.G. (2001). Dissection of nucleophilic and acid-base catalysis in glycosidases. *Current Opinion in Chemical Biology*. **5**, 643-649.

Zubay, G.L. (1998). Biochemistry. Fourth Edition. Wm. C. Brown Publishers.

Zverlov, V.V., Velikodvorskaya, G.A. and Schwarz, W.H. (2003). Two new cellulosome components encoded downstream of *celI* in the genome of *Clostridium thermocellum*: the non-processive endoglucanase CelN and the possibly structural protein CseP. *Microbiology*. **149**, 515-524.

Appendix A

Chemicals, media and enzymes used in this study.

A1: Chemicals

Acros
- Alanine
- Alcian Blue GX8
- Bicine
- 1,4-Butanediol
- Cysteine
- 3,5-Dinitrosalicylic acid
- 1,6-Hexanediol
- Imidazole
- Lithium acetate
- 2-Methyl-2,4-pentanediol
- Phenylalanine
- Serine
- Sodium phosphate (dibasic)
- Sodium phosphate (monobasic)
- Threonine

BDH Laboratory Supplies
- Nitric acid
- Propan-2-ol

Bio-Gene
- 5-bromo-4-chloro-3-indoyl-β-D-galactoside (X-GAL)

Citrus Colloids Ltd
- Apple pectin, experimental blend
- Lime pectin, experimental blend

Fisher BioReagents
- Acrylamide/Bisacrylamide 37.5:1, 40 % solution
- Coomassie Blue R-250
- Lysozyme, egg white
- Methanol
- Phenol

Fisher Chemicals
- Acetic acid, glacial
- Ammonium acetate
- Ammonium citrate
- Ammonium dihydrogen phosphate

Ammonium nitrate
di-Ammonium hydrogen phosphate
tert-Butyl alcohol
Cadmium chloride
Cadmium sulphate
Caesium chloride
Cobalt chloride
Calcium acetate
Dimethylsulphoxide (DMSO)
1,4-Dioxane
Ethyleneglycol
Formaldehyde
Iron (II) chloride
Lithium chloride
Lithium nitrate
Lithium sulphate
Magnesium acetate
Magnesium nitrate
Morpholinoethanesulfonic acid (MES)
PEG 200
PEG 400
PEG 600
PEG 1000
PEG 1500
PEG 4000
PEG 6000
PEG 8000
Polypropyleneglycol 2000
Potassium bromide
Tri-Potassium acetate
Potassium fluoride
Potassium formate
Potassium iodide
Potassium nitrate
Potassium sodium tartrate
Potassium sulphate
Potassium thiocyanate
Sodium cacodylate
Sodium citrate
Sodium iodide
Sodium fluoride
Sodium formate
Sodium nitrate
Sodium sulphate
Sodium tartrate
Sodium thiocyanate
Tryptophan
Zinc acetate
Zinc sulphate

Fluka

- Jeffamine
- Tri-Lithium citrate
- PEG 550mme
- PEG 2000mme
- PEG 5000mme
- PEG 10000
- PEG 20000

Hampton Research

- Aqua Sil

Hayman Ltd

- Absolute alcohol (ethanol)

Megazyme

- Barley β-glucan
- Carboxymethylcellulose
- Carobgalactomannan
- Cellopentaose (G_5)
- Cellohexaose (G_6)
- CM-Pachyman
- Locust bean galactomannan
- Methylcellulose
- Pectin galactan
- Potato galactan
- Rhamnogalacturonan
- Rye arabinoxylan
- Sugar beet arabinan
- Wheat arabinoxylan
- Xylobiose (X_2)
- Xylotriose (X_3)
- Xylotetraose (X_4)
- Xylopentaose (X_5)
- Xylohexaose (X_6)

Melford Laboratories Ltd

- Agarose (High gel strength)
- Glycine
- Isopropyl-β-D-Thiogalactopyranoside (IPTG)
- Kanamycin monosulphate
- Tris [Hydroxymethyl] aminomethane (Tris-HCl)

Riedel-deHaën

- Hydrochloric acid
- Sulphuric acid

Sigma

- Acrylamide/Bisacrylamide 37.5:1, 30 % solution
- Ammonium chloride

Ammonium fluoride
Ammonium formate
Ammonium persulphate (APS)
Ammonium sulphate
Ammonium tartrate
Ampicillin (D[-]-α-Aminobenzylpenicillin)
Arginine
Asparagine
Aspartic acid
Barium chloride, dihydrate
Birchwood xylan
Boric Acid
Bradford reagent
Bovine serum albumin, Fraction V (BSA)
Bromophenol Blue
Calcium chloride, dihydrate
Chloramphenicol
Citric acid, trisodium salt dihydrate
Cobalt chloride, hexahydrate
Collagen, Type I, insoluble, from bovine achilles tendon
Ethidium bromide
Ethylene diamine tetraacetic acid, disodium salt (EDTA)
Ferrous sulphate, heptahydrate
D-(+)-Glucose
Glycerol
Glutamic acid
Glutamine
N-[2-Hydroxyethyl]piperazine-N'-[2-ethanesulphonic acid] (HEPES)
Hexaamminecobalt (III) chloride
Hexadecyltrimethylammonium bromide
Histidine
Hydroxyethyl cellulose
Isoleucine
Isopropanol
Lauryl sulphate, sodium salt (SDS)
Leucine
Lysine
Magnesium chloride, hexahydrate
Magnesium sulphate, heptahydrate
Manganese chloride, tetrahydrate
2-Mercaptoethanol
Methyl mercury (II) chloride
Niacinamide
Nickel (II) chloride, hexahydrate
Nickel sulphate
Oat spelt xylan
PEG 3350
Platinum (II) 2,2',6',2-terpyridine, dihydrate
Polyethyleneimine, (50 % w/v)
Potassium acetate

Potassium chloride
Potassium dichromate
Potassium phosphate, dibasic
Potassium phosphate, monobasic
Proline
Pyridoxine monohydrochloride
Riboflavin
Silver nitrate
Sodium acetate, anhydrous
Sodium carbonate
Sodium chloride
Sodium hydroxide
Sodium sulphite, anhydrous
Sucrose
N,N,N',N'-tetramethylethylene diamine (TEMED)
Tetracycline
Thiamine
Triton X-100
Tyrosine
Uranyl acetate
Urea
Valine
Xylose (X_1)

A2: Media

Oxoid

Agar (Bacteriological agar N° 1)
NZ Amine (Casein hydrolysate)
Tryptone
Yeast Extract

A3: Enzymes

The enzymes, and reaction buffers used in this study are shown in the Tables A3i, A3ii and A3iii.
Invitrogen Corp.:

Enzyme	Buffer composition 1 x
Pfx (DNA polymerase)	Patented

Table A3i: Enzyme and reaction buffer (1 x).

New England Biolabs:

Enzyme	Buffer composition 1 x
Dpn I	Buffer 4 (20 mM Tris-acetate, 10 mM Magnesium acetate, 50 mM Potassium acetate, 1 mM DTT, pH 7.9)
Hind III	Buffer 2 (10 mM Tris-HCl, 10 mM $MgCl_2$, 50 mM NaCl, 1 mM DTT, pH 7.9)
Nde I	Buffer 4 (20 mM Tris-acetate, 10 mM Magnesium acetate, 50 mM Potassium acetate, 1 mM DTT, pH 7.9)
Xho I	Buffer 2 (10 mM Tris-HCl, 10 mM $MgCl_2$, 50 mM NaCl, 1 mM DTT, pH 7.9)
Lysozyme (Hen egg white)	Used un-buffered
T4 DNA ligase	50 mM Tris-HCl pH 7.5, 10 mM $MgCl_2$, 10 mM DTT, 1mM ATP, 25 $\mu g\ ml^{-1}$ acetylated BSA

Table A3ii: Enzymes and reaction buffers (1 x).

Promega:

Enzyme	Buffer composition 1 x
Alkaline phosphatase (Calf intestinal)	50 mM Tris-HCl pH 9.3, 1 mM $MgCl_2$, 0.1 mM $ZnCl_2$, 1 mM spermidine
*Bam*H I	Buffer E (6 mM Tris-HCl, 6 mM $MgCl_2$, 100 mM NaCl, 1 mM DTT, pH 7.5)
*Eco*R I	Buffer H (90 mM Tris-HCl, 10 mM $MgCl_2$, 50 mM NaCl, pH 7.5)
*Eco*R V	Buffer D (6 mM Tris-HCl, 6 mM $MgCl_2$, 150 mM NaCl, 1 mM DTT, pH 7.9)
Sal I	Buffer D (6 mM Tris-HCl, 6 mM $MgCl_2$, 150 mM NaCl, 1 mM DTT, pH 7.9)
T4 DNA ligase	30 mM Tris-HCl pH 7.8, 10 mM $MgCl_2$, 10 mM DTT, 1mM ATP, 5 % PEG
Pfu (DNA polymerase)	10 mM KCl, 10 mM $(NH_4)_2SO_4$, 20 mM Tris-HCl pH 8.8, 2 mM $MgSO_4$, 0.1 % Triton® X-100, 0.1 mg ml^{-1} BSA (nuclease-free)
Taq (DNA polymerase)	10 mM Tris-HCl pH 9.0, 1.5 mM $MgCl_2$, 50mM KCl, 0.1 % Triton® X-100

Table A3iii: Enzymes and reaction buffers (1 x).

A4: Size standards

DNA

New England Biolabs:

1 Kb DNA ladder (10, 8, 6, 5, 4, 3, 2, 1.5, 1, 0.5 Kb)

Promega:

100 bp DNA ladder (1500, 1000, 900, 800, 700, 600, 500, 400, 300, 200, 100 bp)

Lambda DNA / *Hin*d III ladder (23130, 9416, 6557, 4361, 2322, 2027, 564, 125 bp).

Protein

Sigma:

Proteins	M.W. (Da)	High	Low
Myosin, rabbit muscle	205000	X	
β-Galactosidase, *E. coli*	116000	X	
Phosphorylase b, rabbit muscle	97000	X	
Fructose-6-phosphate Kinase, rabbit muscle	84000	X	
Albumin, bovine serum	66000	X	X
Glutamic Dehydrogenase, bovine liver	55000	X	
Ovalbumin, chicken egg	45000	X	X
Glyceraldehydes-3-phosphate Dehydrogenase, rabbit muscle	36000	X	X
Carbonic Anhydrase, bovine erythrocytes	29000		X
Trypsinogen, bovine pancreas	24000		X
Trypsin Inhibitor, soybean	20000		X
α-Lactalbumin, bovine milk	14200		X
Aprotinin, bovine lung	6500		X

Table A4: Molecular weight distribution in Sigma Markers.

A5: Kits

Macherey-Nagel GmbH (Germany):
NucleoSpin® Extract 2 in 1
NucleoSpin® Plasmid

NucleoSpin® Extract / Plasmid kit buffers	Buffer components / function
Buffer A1	Resuspension buffer: contains RNase
Buffer A2	Lysis buffer
Buffer A3	Binding buffer: contains guanidinium hydrochloride
Buffer A4	Wash buffer: contains ethanol
Buffer AE	Elution buffer: 5 mM Tris-HCl pH 8.5
Buffer AW	Wash buffer: guanidinium hydrochloride
Buffer NT1	Binding buffer
Buffer NT2	
Buffer NT3	Contains ethanol
Buffer NE	Elution buffer: 5 mM Tris-HCl, pH 8.5

Table A5i: Components of NucleoSpin® kits.
Absolute components are patented.

Stratagene:
QuikChange™ Site-Directed Mutagenesis Kit

Materials provided
Pfu turbo DNA polymerase (2.5 U μl^{-1})
Reaction buffer (10 x)
Dpn I restriction enzyme (10 U μl^{-1})
dNTP mix

Table A5ii: Components provided with QuikChange™ Site-Directed Mutagenesis kit.

A6: Resins

Amersham Pharmacia Biotech AB
Chelating Sepharose™ fast flow resin
Sephadex™ G-25 medium resin

Clontech Labs Inc.
Talon™ resin

Appendix B

Vectors used in this study and their expanded polylinker region.

B1: pGEM®-T Easy (Promega)

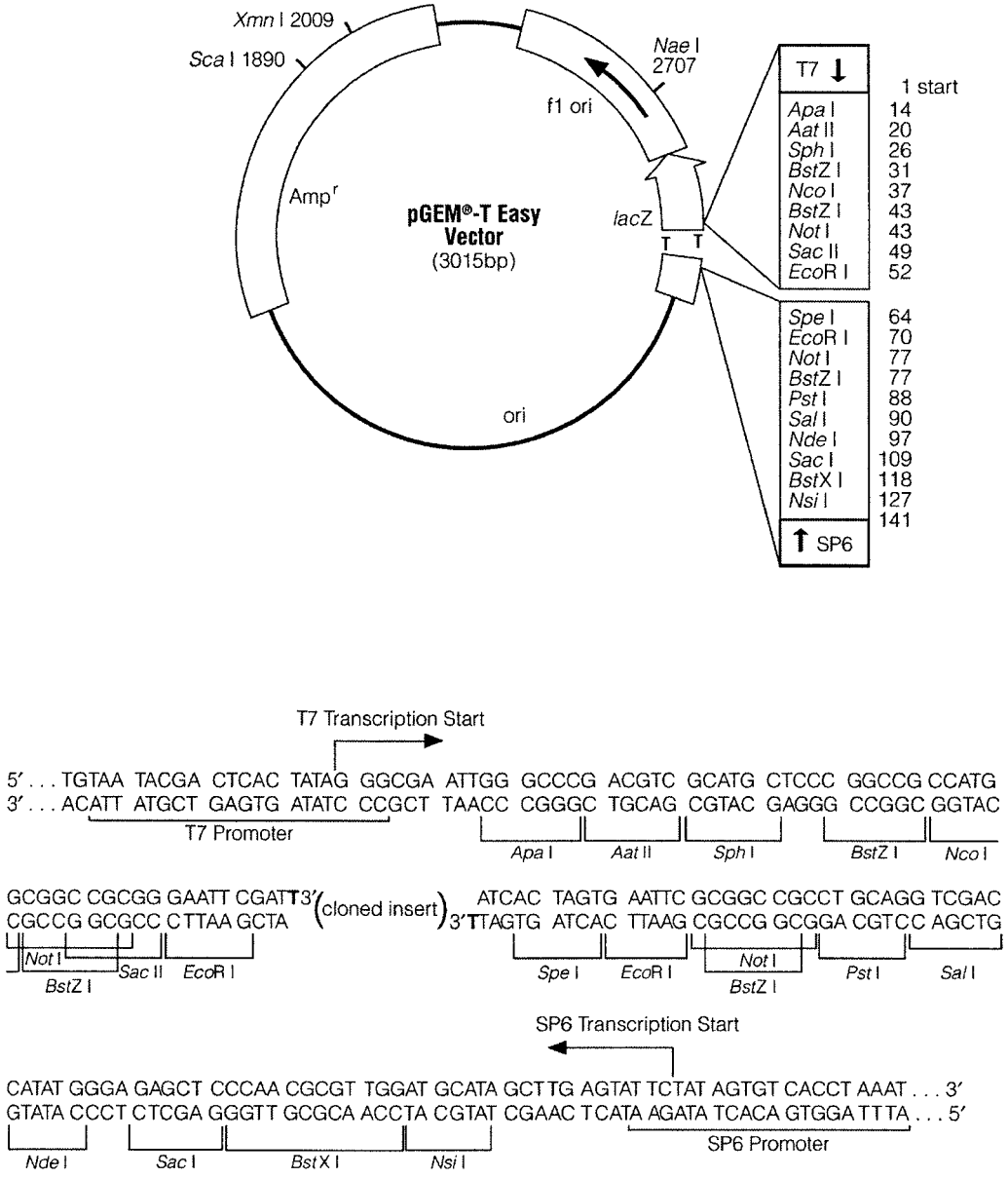


Figure B1: pGEM®-T Easy vector.
Plasmid DNA prepared by digestion with *EcoR* V and 3' terminal thymidines added.

B2: pCR®-Blunt (Invitrogen Corp.)

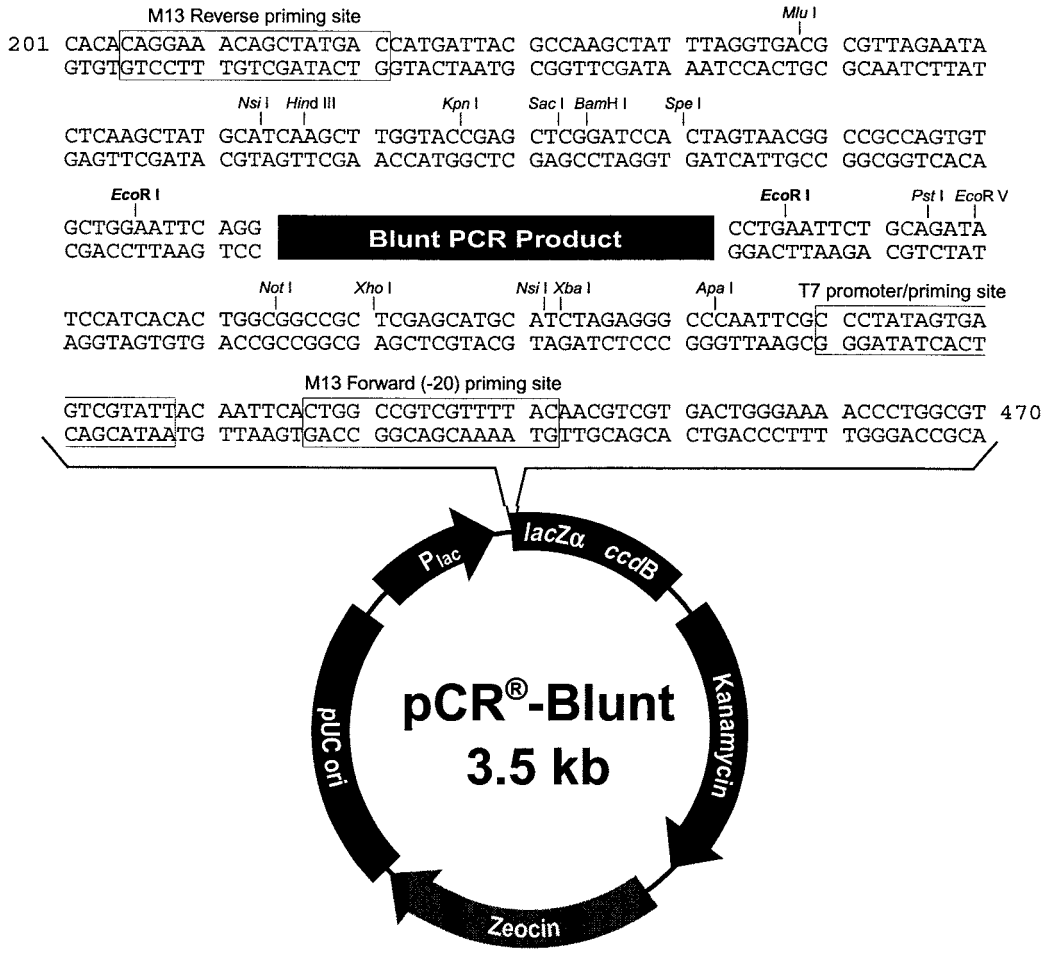
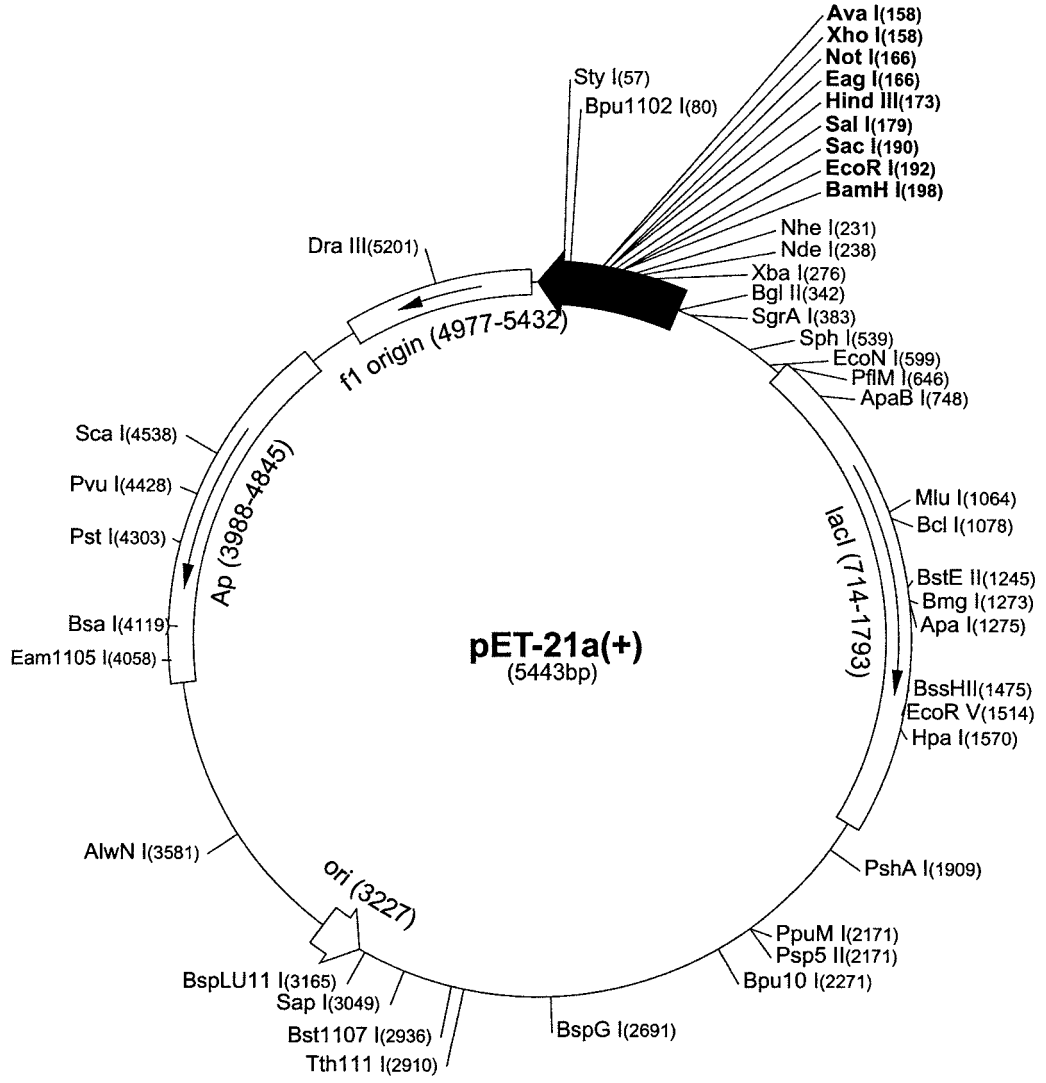


Figure B2: pCR®-Blunt vector.

B3: pET-21a (Novagen)



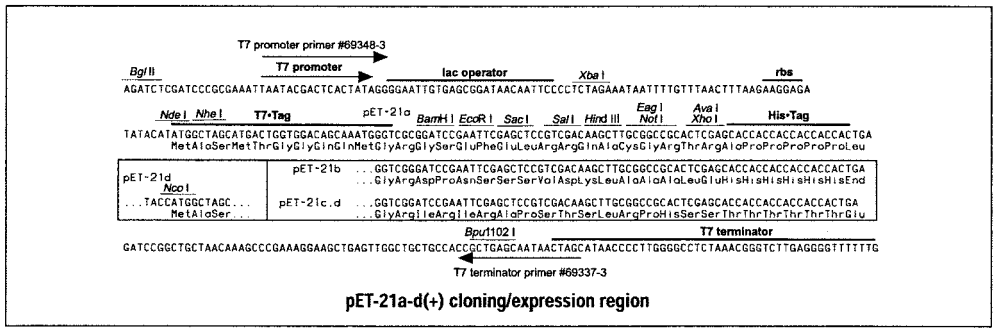
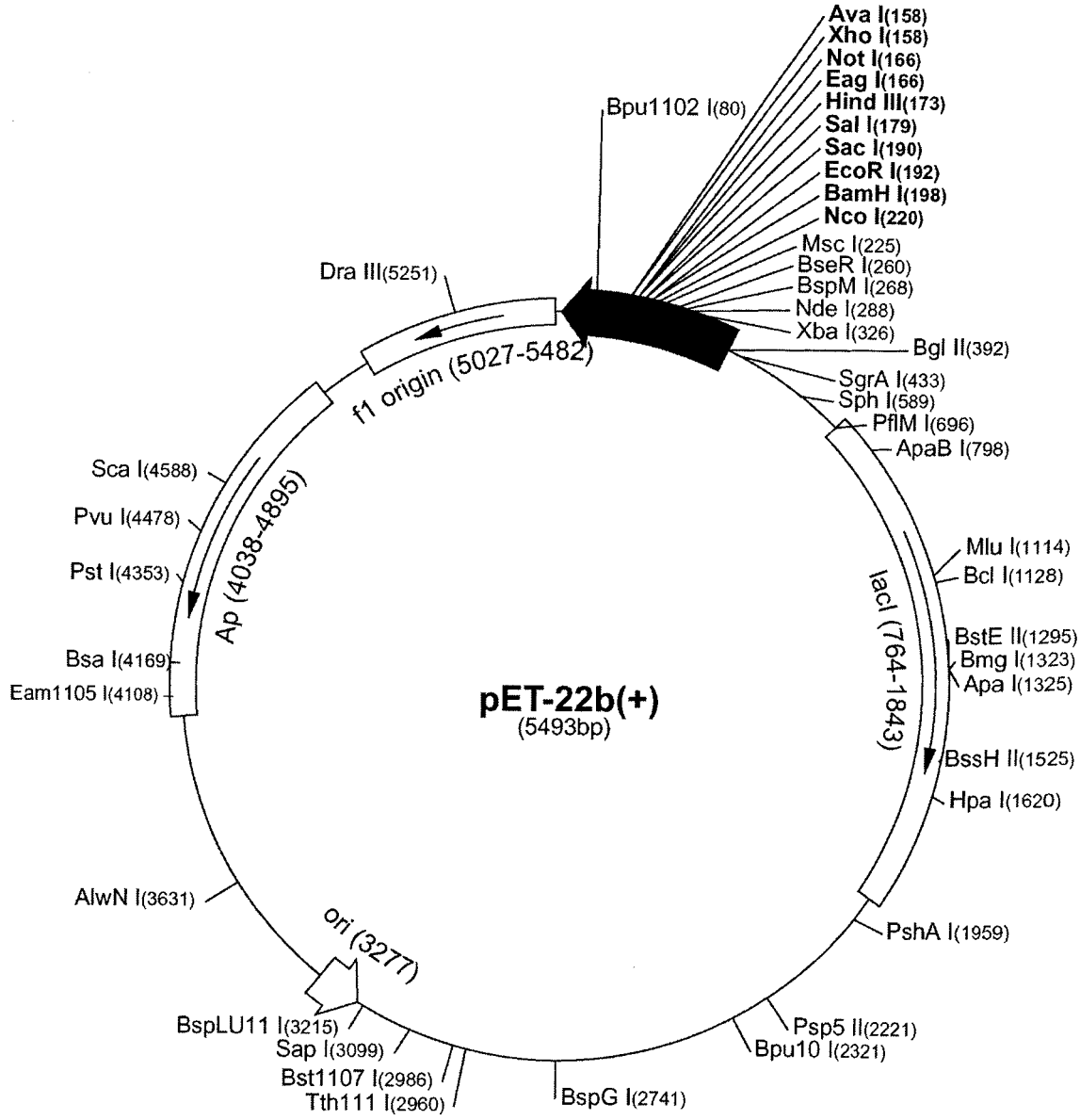


Figure B3: pET-21a vector.

B4: pET-22b (Novagen)



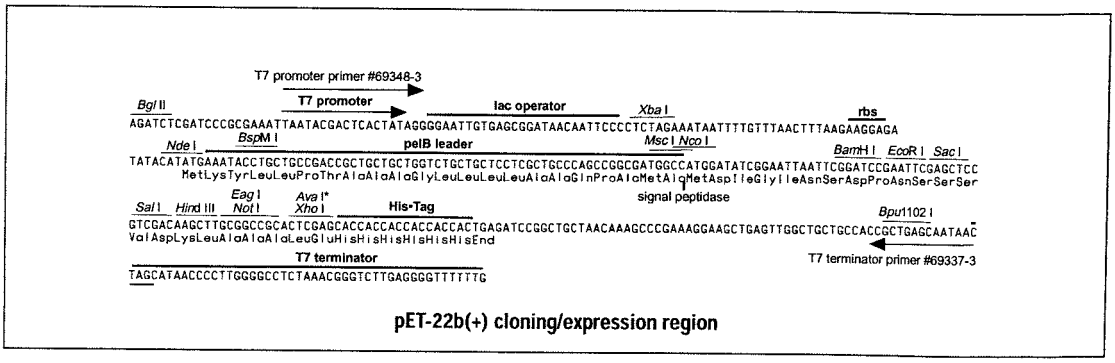
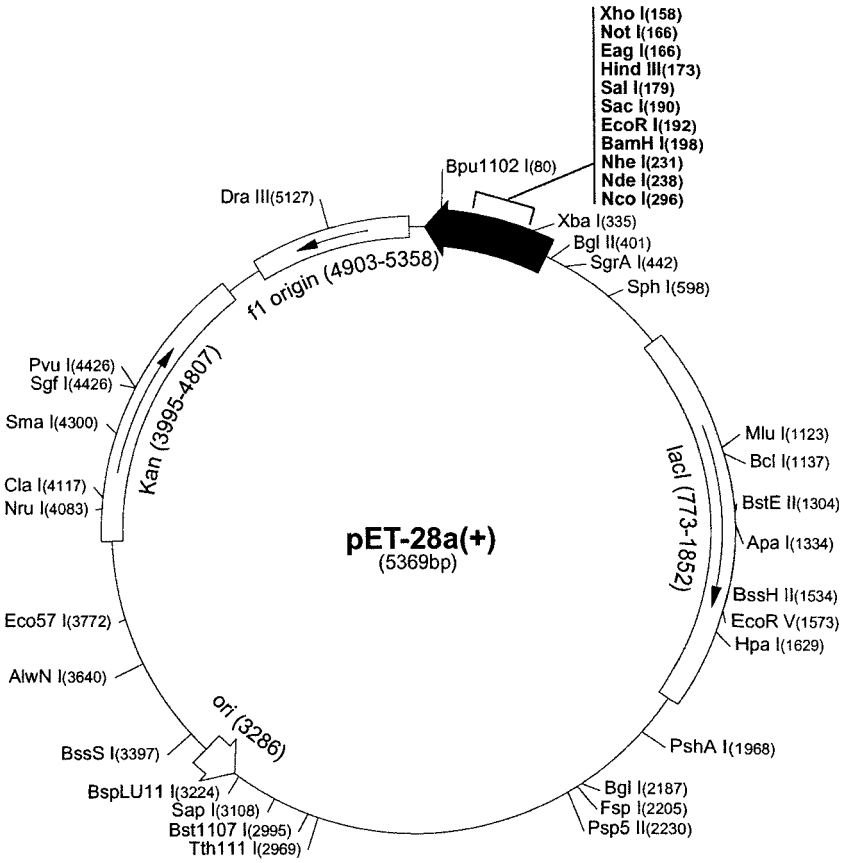


Figure B4: pET-22b vector.

B5: pET-28a (Novagen)



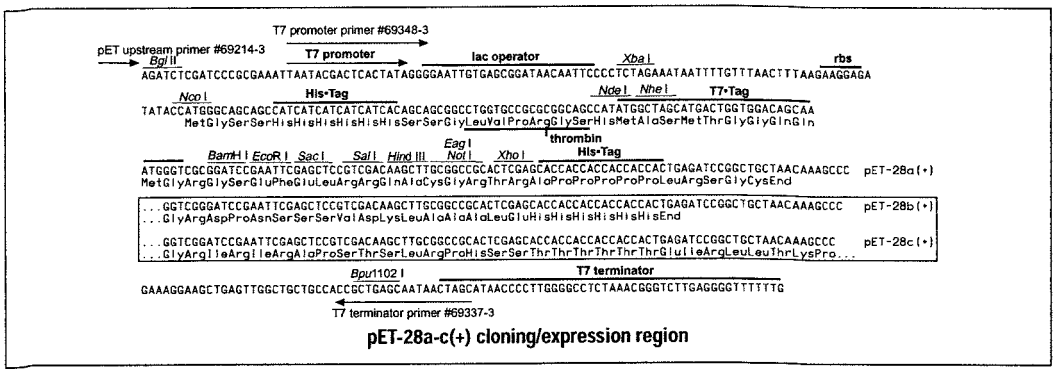


Figure B5: pET-28a vector.

Appendix C

Protein crystallisation screens

Tables CI, CII, CIII, CIV and CV show the solutions used to screen for crystal development. Those screens in tables CI and CII are named Clear Strategy Screen (CSS) I and II respectively from Molecular Dimensions. The following abbreviations describe precipitants: polyethylene glycol (PEG), monomethylether (MME), Jeffamine (O-(2-Aminopropyl)-O'-(2-methoxy-ethyl)polypropylene glycol 500).

C1: CSS 1

Tube Number	Salt	Precipitant
1	0.3M Sodium Acetate	25% (w/v) PEG 2Kmme
2	0.2M Lithium Sulphate	25% (w/v) PEG 2Kmme
3	0.2M Magnesium Chloride	25% (w/v) PEG 2Kmme
4	0.2M Potassium Bromide	25% (w/v) PEG 2Kmme
5	0.2M Potassium Thiocyanate	25% (w/v) PEG 2Kmme
6	0.8M Sodium Formate	25% (w/v) PEG 2Kmme
7	0.3M Sodium Acetate	15% (w/v) PEG 4K
8	0.2M Lithium Sulphate	15% (w/v) PEG 4K
9	0.2M Magnesium Chloride	15% (w/v) PEG 4K
10	0.2M Potassium Bromide	15% (w/v) PEG 4K
11	0.2M Potassium Thiocyanate	15% (w/v) PEG 4K
12	0.8M Sodium Formate	15% (w/v) PEG 4K
13	0.3M Sodium Acetate	10% (w/v) PEG 8K, 10% (w/v) PEG 1K
14	0.2M Lithium Sulphate	10% (w/v) PEG 8K, 10% (w/v) PEG 1K
15	0.2M Magnesium Chloride	10% (w/v) PEG 8K, 10% (w/v) PEG 1K
16	0.2M Potassium Bromide	10% (w/v) PEG 8K, 10% (w/v) PEG 1K
17	0.2M Potassium Thiocyanate	10% (w/v) PEG 8K, 10% (w/v) PEG 1K
18	0.8M Sodium Formate	10% (w/v) PEG 8K, 10% (w/v) PEG 1K
19	0.3M Sodium Acetate	8% (w/v) PEG 20K, 8% (w/v) PEG 550mme
20	0.2M Lithium Sulphate	8% (w/v) PEG 20K, 8% (w/v) PEG 550mme
21	0.2M Magnesium Chloride	8% (w/v) PEG 20K, 8% (w/v) PEG 550mme
22	0.2M Potassium Bromide	8% (w/v) PEG 20K, 8% (w/v) PEG 550mme
23	0.2M Potassium Thiocyanate	8% (w/v) PEG 20K, 8% (w/v) PEG 550mme
24	0.8M Sodium Formate	8% (w/v) PEG 20K, 8% (w/v) PEG 550mme

Table CI: CSS I components.

C2: CSS II

Tube Number	Salt	Precipitant
1	1.5M Ammonium Sulphate	
2	0.8M Lithium Sulphate	
3	2M Sodium Formate	
4	0.5M Potassium dihydrogen Phosphate	
5	0.2M Calcium Acetate	25% (w/v) PEG 2Kmme
6	0.2M Calcium Acetate	15% (w/v) PEG 4K
7	2.7M Ammonium Sulphate	
8	1.8M Lithium Sulphate	
9	4M Sodium Formate	
10	1M Potassium dihydrogen Phosphate	
11	0.2M Calcium Acetate	10% (w/v) PEG 8K, 10% (w/v) PEG 1K
12	0.2M Calcium Acetate	8% (w/v) PEG 20K, 8% (w/v) PEG 550mme
13		40% (v/v) MPD
14		40% (v/v) 2,3-Butane-diol
15	5mM Cadmium Chloride	20% (w/v) PEG 4K
16	0.15M Potassium Thiocyanate	20% (w/v) PEG 550mme
17	0.15M Potassium Thiocyanate	20% (w/v) PEG 600
18	0.15M Potassium Thiocyanate	20% (w/v) PEG 1500
19		35% (v/v) Isopropanol
20		30% (v/v) Jeffamine 600M pH 6.5
21	5mM Nickel Chloride	20% (w/v) PEG 4K
22	0.15M Potassium Thiocyanate	18% (w/v) PEG 3350
23	0.15M Potassium Thiocyanate	18% (w/v) PEG 5Kmme
24	0.15M Potassium Thiocyanate	15% (w/v) PEG 6K

Table CII: CSS II components.

C3: Hampton screen I

Tube Number	Salt	Buffer	Precipitant
1	0.02M Calcium Chloride dihydrate	0.1M Sodium Acetate trihydrate pH 4.6	30% (v/v) 2-Methyl-2, 4-pentanediol
2			0.4M Potassium Sodium Tartrate tetrahydrate
3			0.4M Ammonium dihydrogen Phosphate
4		0.1M Tris Hydrochloride pH 8.5	2M Ammonium Sulphate
5	0.2M tri-Sodium Citrate dihydrate	0.1M HEPES-Sodium pH 7.5	30% (v/v) 2-Methyl-2, 4-pentanediol
6	0.2M Magnesium Chloride hexahydrate	0.1M Tris Hydrochloride pH 8.5	30% (w/v) PEG 4000
7		0.1M Sodium Cacodylate pH 6.5	1.4M Sodium Acetate trihydrate
8	0.2M tri-Sodium Citrate dihydrate	0.1M Sodium Cacodylate pH 6.5	30% (v/v) iso-Propanol
9	0.2M Ammonium Acetate	0.1M tri-Sodium Citrate dihydrate pH 5.6	30% (w/v) PEG 4000
10	0.2M Ammonium Acetate	0.1M Sodium Acetate trihydrate pH 4.6	30% (w/v) PEG 4000
11		0.1M tri-Sodium Citrate dihydrate pH 5.6	1M Ammonium dihydrogen Phosphate
12	0.2M Magnesium Chloride hexahydrate	0.1M HEPES-Sodium pH 7.5	30% (v/v) iso-Propanol
13	0.2M tri-Sodium Citrate dihydrate	0.1M Tris Hydrochloride pH 8.5	30% (v/v) PEG 400
14	0.2M Calcium Chloride dihydrate	0.1M HEPES-Sodium pH 7.5	28% (v/v) PEG 400
15	0.2M Ammonium Sulphate	0.1M Sodium Cacodylate pH 6.5	30% (w/v) PEG 8000
16		0.1M HEPES-Sodium pH 7.5	1.5M Lithium Sulphate monohydrate
17	0.2M Lithium	0.1M Tris	30% (w/v) PEG

	Sulphate monohydrate	Hydrochloride pH 8.5	4000
18	0.2M Magnesium Acetate tetrahydrate	0.1M Sodium Cacodylate pH 6.5	20% (w/v) PEG 8000
19	0.2M Ammonium Acetate	0.1M Tris Hydrochloride pH 8.5	30% (v/v) iso-Propanol
20	0.2M Ammonium Sulphate	0.1M Sodium Acetate trihydrate pH 4.6	25% (w/v) PEG 4000
21	0.2M Magnesium Acetate tetrahydrate	0.1M Sodium Cacodylate pH 6.5	30% (v/v) 2-Methyl-2, 4-pentanediol
22	0.2M Sodium Acetate trihydrate	0.1M Tris Hydrochloride pH 8.5	30% (w/v) PEG 4000
23	0.2M Magnesium Chloride hexahydrate	0.1M HEPES-Sodium pH 7.5	30% (v/v) PEG 400
24	0.2M Calcium Chloride dihydrate	0.1M Sodium Acetate trihydrate pH 4.6	20% (v/v) iso-Propanol
25		0.1M Imidazole pH 6.5	1M Sodium Acetate trihydrate
26	0.2M Ammonium Acetate	0.1M tri-Sodium Citrate dihydrate pH 5.6	30% (v/v) 2-Methyl-2, 4-pentanediol
27	0.2M tri-Sodium Citrate dehydrate	0.1M HEPES-Sodium pH 7.5	20% (v/v) iso-Propanol
28	0.2M Sodium Acetate trihydrate	0.1M Sodium Cacodylate pH 6.5	30% (w/v) PEG 8000
29		0.1M HEPES-Sodium pH 7.5	0.8M Potassium Sodium Tartrate tetrahydrate
30	0.2M Ammonium Sulphate		30% (w/v) PEG 8000
31	0.2M Ammonium Sulphate		30% (w/v) PEG 4000
32			2M Ammonium Sulphate
33			4M Sodium Formate
34		0.1M Sodium Acetate trihydrate pH 4.6	2M Sodium Formate
35		0.1M HEPES-Sodium pH 7.5	0.8M Sodium dihydrogen Phosphate, 0.8M Potassium

			dihydrogen Phosphate
36		0.1M Tris Hydrochloride pH 8.5	8% (w/v) PEG 8000
37		0.1M Sodium Acetate trihydrate pH 4.6	8% (w/v) PEG 4000
38		0.1M HEPES-Sodium pH 7.5	1.4M tri-Sodium Citrate dihydrate
39		0.1M HEPES-Sodium pH 7.5	2% (v/v) PEG 400, 2M Ammonium Sulphate
40		0.1M tri-Sodium Citrate dihydrate pH 5.6	20% (v/v) iso-Propanol, 20% (w/v) PEG 4000
41		0.1M HEPES-Sodium pH 7.5	10% (v/v) iso-Propanol, 20% (w/v) PEG 4000
42	0.05M Potassium dihydrogen Phosphate		20% (w/v) PEG 8000
43			30% (w/v) PEG 1500
44			0.2M Magnesium Formate
45	0.2M Zinc Acetate dihydrate	0.1M Sodium Cacodylate pH 6.5	18% (w/v) PEG 8000
46	0.2M Calcium Acetate hydrate	0.1M Sodium Cacodylate pH 6.5	18% (w/v) PEG 8000
47		0.1M Sodium Acetate trihydrate pH 4.6	2M Ammonium Sulphate
48		0.1M Tris Hydrochloride pH 8.5	2M Ammonium dihydrogen Phosphate

Table CIII: Hampton screen I components.

C4: Hampton screen II

Tube Number	Salt	Buffer	Precipitant
1	2M Sodium Chloride		10% (w/v) PEG 6000
2	0.5M Sodium Chloride; 0.1M Magnesium Chloride hexahydrate		0.01M Hexadecyltrimethylammonium Bromide
3			25% (v/v) Ethylene Glycol
4			35% (v/v) Dioxane
5	2M Ammonium Sulphate		5% (v/v) iso-Propanol
6			1M Imidazole pH 7.0
7			10% (w/v) PEG 1000; 10% (w/v) PEG 8000
8	1.5M Sodium Chloride		10% (v/v) Ethanol
9		0.1M Sodium Acetate trihydrate pH 4.6	2M Sodium Chloride
10	0.2M Sodium Chloride	0.1M Sodium Acetate trihydrate pH 4.6	30% (v/v) MPD
11	0.01M Cobaltous Chloride hexahydrate	0.1M Sodium Acetate trihydrate pH 4.6	1M 1,6 Hexanediol
12	0.1M CadmiumChloride dihydrate	0.1M Sodium Acetate trihydrate pH 4.6	30% (v/v) PEG 400
13	0.2M Ammonium Sulphate	0.1M Sodium Acetate trihydrate pH 4.6	30% (w/v) PEG 2000mme
14	0.2M Potassium Sodium Tartrate tetrahydrate	0.1M tri-Sodium Citrate dihydrate pH 5.6	2M Ammonium Sulphate
15	0.5M Ammonium Sulphate	0.1M tri-Sodium Citrate dihydrate pH	1M Lithium Sulphate monohydrate

		5.6	
16	0.5M Sodium Chloride	0.1M tri-Sodium Citrate dihydrate pH 5.6	2% (w/v) Ethylene Imine Polymer
17		0.1M tri-Sodium Citrate dihydrate pH 5.6	35% (v/v) tert-Butanol
18	0.01M Ferric Chloride hexahydrate	0.1M tri-Sodium Citrate dihydrate pH 5.6	10% (v/v) Jeffamine M-600
19		0.1M tri-Sodium Citrate dihydrate pH 5.6	2.5M 1,6 Hexanediol
20		0.1M MES pH 6.5	1.6M Magnesium Sulphate heptahydrate
21	0.1M Sodium dihydrogen Phosphate; 0.1M Potassium dihydrogen Phosphate	0.1M MES pH 6.5	2M Sodium Chloride
22		0.1M MES pH 6.5	12% (w/v) PEG 20000
23	1.6M Ammonium Sulphate	0.1M MES pH 6.5	10% (v/v) Dioxane
24	0.05M Caesium Chloride	0.1M MES pH 6.5	30% (v/v) Jeffamine M-600
25	0.01M Cobaltous Chloride hexahydrate	0.1M MES pH 6.5	1.8M Ammonium Sulphate
26	0.2M Ammonium Sulphate	0.1M MES pH 6.5	30% (w/v) PEG 5000mme
27	0.01M Zinc Sulphate heptahydrate	0.1M MES pH 6.5	25% (v/v) PEG 550mme
28			1.6M tri-Sodium Citrate dihydrate pH 6.5
29	0.5M Ammonium Sulphate	0.1M HEPES pH 7.5	30% (v/v) MPD
30		0.1M HEPES pH 7.5	10% (w/v) PEG 6000; 5% (v/v) MPD

31		0.1M HEPES pH 7.5	20% (v/v) Jeffamine M-600
32	0.1M Sodium Chloride	0.1M HEPES pH 7.5	1.6M Ammonium Sulphate
33		0.1M HEPES pH 7.5	2M Ammonium Formate
34	0.05M Cadmium Sulphate hydrate	0.1M HEPES pH 7.5	1M Sodium Acetate trihydrate
35		0.1M HEPES pH 7.5	70% (v/v) MPD
36		0.1M HEPES pH 7.5	4.3M Sodium Chloride
37		0.1M HEPES pH 7.5	10% (w/v) PEG 8000; 8% (v/v) Ethylene Glycol
38		0.1M HEPES pH 7.5	20% (w/v) PEG 10000
39	0.2M Magnesium Chloride hexahydrate	0.1M Tris pH 8.5	3.4M 1,6 Hexanediol
40		0.1M Tris pH 8.5	25% (v/v) tert-Butanol
41	0.01M Nickel (II) Chloride hexahydrate	0.1M Tris pH 8.5	1M Lithium Sulphate monohydrate
42	1.5M Ammonium Sulphate	0.1M Tris pH 8.5	12% (v/v) Glycerol anhydrous
43	0.2M Ammonium dihydrogen Phosphate	0.1M Tris pH 8.5	50% (v/v) MPD
44		0.1M Tris pH 8.5	20% (v/v) Ethanol
45	0.01M Nickel (II) Chloride hexahydrate	0.1M Tris pH 8.5	20% (w/v) PEG 2000mme
46	0.1M Sodium Chloride	0.1M Bicine pH 9.0	20% (w/v) PEG 550mme
47		0.1M Bicine pH 9.0	2M Magnesium Chloride hexahydrate
48		0.1M Bicine pH 9.0	2% (v/v) Dioxane; 20% (w/v) PEG 2000mme 10% (w/v) PEG 20000

Table CIV: Hampton screen II components.

C5: PEG ion screen

Tube Number	Salt	Precipitant
1	0.2M Sodium Fluoride	20% (w/v) PEG 3350
2	0.2M Potassium Fluoride	20% (w/v) PEG 3350
3	0.2M Ammonium Fluoride	20% (w/v) PEG 3350
4	0.2M Lithium Chloride anhydrous	20% (w/v) PEG 3350
5	0.2M Magnesium Chloride hexahydrate	20% (w/v) PEG 3350
6	0.2M Sodium Chloride	20% (w/v) PEG 3350
7	0.2M Calcium Chloride dihydrate	20% (w/v) PEG 3350
8	0.2M Potassium Chloride	20% (w/v) PEG 3350
9	0.2M Ammonium Chloride	20% (w/v) PEG 3350
10	0.2M Sodium Iodide	20% (w/v) PEG 3350
11	0.2M Potassium Iodide	20% (w/v) PEG 3350
12	0.2M Ammonium Iodide	20% (w/v) PEG 3350
13	0.2M Sodium Thiocyanate	20% (w/v) PEG 3350
14	0.2M Potassium Thiocyanate	20% (w/v) PEG 3350
15	0.2M Lithium Nitrate	20% (w/v) PEG 3350
16	0.2M Magnesium Nitrate hexahydrate	20% (w/v) PEG 3350
17	0.2M Sodium Nitrate	20% (w/v) PEG 3350
18	0.2M Potassium Nitrate	20% (w/v) PEG 3350
19	0.2M Ammonium Nitrate	20% (w/v) PEG 3350
20	0.2M Magnesium Formate	20% (w/v) PEG 3350
21	0.2M Sodium Formate	20% (w/v) PEG 3350
22	0.2M Potassium Formate	20% (w/v) PEG 3350
23	0.2M Ammonium Formate	20% (w/v) PEG 3350
24	0.2M Lithium Acetate dihydrate	20% (w/v) PEG 3350
25	0.2M Magnesium Acetate tetrahydrate	20% (w/v) PEG 3350
26	0.2M Zinc Acetate dihydrate	20% (w/v) PEG 3350
27	0.2M Sodium Acetate trihydrate	20% (w/v) PEG 3350
28	0.2M Calcium Acetate hydrate	20% (w/v) PEG 3350
29	0.2M Potassium Acetate	20% (w/v) PEG 3350
30	0.2M Ammonium Acetate	20% (w/v) PEG 3350
31	0.2M Lithium Sulphate monohydrate	20% (w/v) PEG 3350
32	0.2M Magnesium Sulphate heptahydrate	20% (w/v) PEG 3350
33	0.2M Sodium Sulphate decahydrate	20% (w/v) PEG 3350
34	0.2M Potassium Sulphate	20% (w/v) PEG 3350

35	0.2M Ammonium Sulphate	20% (w/v) PEG 3350
36	0.2M di-Sodium tartrate dihydrate	20% (w/v) PEG 3350
37	0.2M Potassium Sodium Tartrate tetrahydrate	20% (w/v) PEG 3350
38	0.2M di-Ammonium Tartrate	20% (w/v) PEG 3350
39	0.2M Sodium dihydrogen Phosphate monohydrate	20% (w/v) PEG 3350
40	0.2M di-Sodium hydrogen Phosphate dihydrate	20% (w/v) PEG 3350
41	0.2M Potassium dihydrogen Phosphate	20% (w/v) PEG 3350
42	0.2M di-Potassium hydrogen Phosphate	20% (w/v) PEG 3350
43	0.2M Ammonium dihydrogen Phosphate	20% (w/v) PEG 3350
44	0.2M di-Ammonium hydrogen Phosphate	20% (w/v) PEG 3350
45	0.2M tri-Lithium Citrate tetrahydrate	20% (w/v) PEG 3350
46	0.2M tri-Sodium Citrate dihydrate	20% (w/v) PEG 3350
47	0.2M tri-Potassium Citrate monohydrate	20% (w/v) PEG 3350
48	0.2M di-Ammonium hydrogen Citrate	20% (w/v) PEG 3350

Table CV: PEG ion screen components.

Appendix D

Equipment

D1: Autoclaving

Autoclave sterilisation was achieved using a benchtop Prestige® Medical 2100 Classic autoclave at 121 °C, 32 lb/inch^2 pressure for 20 min.

D2: Incubators

For the growth of bacteria in liquid cultures a GallenKamp orbital shaker was used. Growth of bacteria on solid media was performed in a static GallenKamp incubator. Crystal growth was performed in a LMS cooled incubator.

D3: Freeze drier

Proteins and substrates were lyophilised using the Christ® Alpha 1-2 freeze-drier.

D4: pH meter

All adjustments to the pH of solutions and media were achieved using a Jenway Ion Meter 3340 calibrated with buffers at pH 4.0, 7.0, and 9.2.

D5: Reaction vessels

Unless stated otherwise, volumes up to 0.2 ml were contained in 0.2 ml micro-centrifuge tubes (Starstedt), 0.2-1.5 ml volumes in 1.5 ml micro-centrifuge tubes (Starstedt) and for volumes > 1.5 ml, 28 ml sterile plastic universal containers were used (Fisherbrand®).

D6: Centrifugation

Centrifugation of volumes below 1.5ml was achieved using a small Sigma* 1-15 bench top micro-centrifuge. Volumes above 1.5ml were centrifuged in a large Sigma* 3K18C refrigerated bench top centrifuge, using the rotors and inserts appropriate to the application (Table D4).

Rotor	Vessel containing solution	Insert	RCF	Application
11133	25ml Plastic universal	Clay universal holder (part N° 17049)	2000	Concentration of cells in preparation for chemical competency
11133	30kDa cut-off protein concentrator (Amicon)	Clay universal holder (part N°17049)	2500	Concentration of protein
11133	200ml Plastic centrifuge vessel (part N° 15202)	None	5500	Concentration of cells
12158	30ml Plastic centrifuge tube CEN9534 (part N° 3119-0030)	None	24000	Separation of soluble cell extract, from lysed cells
12131	1.5ml micro-centrifuge tube	None	24000	Isolation of pellet on crude plasmid extraction

Table D6: Centrifuge rotors, vessels and applications.

D7: Agarose gel kits

Electrophoresis of DNA was achieved using an Owl Scientific EASY-CAST™ electrophoresis system powered by an E-C 570-90 E-C Apparatus Corporation power pack.

D8: SDS-PAGE gel kit

Electrophoresis of proteins was achieved using a Bio-Rad Mini-PROTEAN® 3 Cell kit powered by an E-C 570-90 E-C apparatus corporation power pack.

D9: Gel documentation

Visualisation of agarose and SDS-PAGE gels was achieved using a Bio-Rad Gel Doc 2000 system and Quantity One™ software. Hard copies of the gel picture were produced using a Mitsubishi Video Copy Processor (Model P91), with Mitsubishi thermal paper (K65HM-$_{CE}$/High density type, 110 mm x 21 m).

D10: Sonication

Cell lysis was achieved using a MSE Soniprep 150 ultra-sonication machine.

D11: Large scale protein purification

For the large scale purification of proteins a FPLC® (Applied Biosystems) system was used equipped with the hardware shown in Table D9, together with a EYELA microtube Pump MP-3 peristaltic pump.

Hardware (AB)	Part Number or code
Pumps x 2	P-500
Conductivity Meter	Conductivity Monitor Code N° 18-1500-00
UV visible Spectrophotometer	LKB- Uvicord VW 2251
Fraction collector	FRAC – 100

Table D11: Hardware used for large-scale protein purification.

D12: Spectrophotometer

Spectrophotometric measurements were made using a UV visible Heλios α Spectronic Unicam spectrophotometer.

D13: PCR machine

All PCR reactions were performed in an Eppendorf MasterCycler™ PCR machine.

D14: Microtitre plate reader

All DNSA assays were measured A_{570} using a FL600 FAK microtitre plate reader (Bio-Tek Kontron Instruments Inc.) using MicroTek OS software (MTK version 2.5).